

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 4, 2002, 08:56:12 ; Search time 34.94 Seconds
(without alignments)
2384.240 Million cell updates/sec

Title: US-09-810-186-1
Perfect score: 3996
Sequence: 1 MTVPKISVNDGKLVHGKT.....TFSIPVSEEMYSIEIOW 750

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq.033802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3996	100.0	750	20	AA17418 Soybean raffinose
2	2979.5	74.6	751	21	AAV70980 Wheat raffinose sy
3	2874	71.9	756	21	AAV70975 Corn raffinose syn
4	2827.5	70.8	763	21	AAV70977 Rice raffinose syn
5	2117.5	53.0	770	21	AAV70976 Rice raffinose syn
6	2110.5	52.8	841	21	AAV70981 Wheat raffinose sy
7	1590	39.8	508	21	AAV70974 Corn raffinose syn
8	1430.5	35.8	758	21	AAV70978 Soybean raffinose
9	1415.5	35.4	777	20	AAV70978 Mustard raffinose
10	1394	34.9	784	19	AAW53570 Cucurbit raffinose
11	1394	34.9	784	20	AAV17417 Cucurbit raffinose

12	1385	34.7	781	19	AAW57887
13	1385	34.7	781	20	AAV30143
14	1385	34.7	781	22	AAW49400
15	1379	34.5	783	19	AAV32073
16	1370.5	34.3	799	19	AAW57886
17	1370.5	34.3	799	20	AAV30142
18	1367.5	34.2	780	22	AAW98659
19	1117.5	28.0	587	19	AAW57888
20	1117.5	28.0	587	20	AAV30144
21	1095.5	27.4	572	20	AAV32075
22	538.5	13.5	229	21	AAW28628
23	484.5	12.1	265	20	AAV32072
24	477	11.9	271	19	AAW57889
25	477	11.9	271	20	AAV30145
26	376.5	9.4	193	21	AAW28629
27	376.5	9.4	193	21	AAW28630
28	376.5	9.4	193	21	AAW28630
29	178	4.5	45	21	AAV70979
30	178	4.5	87	21	AAW15267
31	120.5	3.0	157	21	AAW15266
32	120.5	3.0	2122	22	ABG12590
33	120.5	3.0	2652	22	ABG11747
34	120.5	3.0	2657	22	ABG04691
35	119.5	3.0	920	22	AAV97652
36	117.5	2.9	1074	19	AAW64186
37	117.5	2.9	4544	15	AAW47861
38	117.5	2.9	4544	15	AAW60517
39	117.5	2.9	4544	22	AAW79091
40	116	2.9	4563	22	ABW13533
41	115	2.9	1732	22	ABW71861
42	115	2.9	478	19	AAW64173
43	115	2.9	914	20	AAW86429
44	111	2.8	928	20	AAW88418
45	109.5	2.7	2209	3	AAW20037
				15	AAW62487

ALIGNMENTS

RESULT	1
AAV17418	
ID	AAV17418 standard; Protein; 750 AA.
XX	
AC	AAV17418;
XX	
DE	29-JUL-1999 (first entry)
XX	
DE	Soybean raffinose synthase.
XX	
KW	Raffinose synthase; sucrose; galactinol.
XX	
OS	Glycine max.
XX	
PN	JPI1123080-A.
XX	
PD	11-MAY-1999.
XX	
PF	24-OCT-1997; 97JP-0292969.
XX	
PR	24-OCT-1997; 97JP-0292969.
XX	
PA	(AJIN) AJINOMOTO KK.
XX	
DR	WPI; 1999-340516/29.
XX	
DR	N-PSDB; AAW61239.
XX	
PT	New raffinose synthase gene - for production of raffinose from
XX	sucrose and galactinol
PS	Claim 2; Page 32-34; 37pp; Japanese.
XX	
CC	The present invention describes a raffinose synthase, having an actively
CC	forming raffinose from sucrose and galactinol. The raffinose synthase

Soybean raffinose
Amino acid sequenc
Soybean raffinose
Sugarcane raffinose
Broad bean raffin
Amino acid sequenc
Soybean protein: S
Japanese artichoke
Amino acid sequenc
Rapeseed raffinose
Arabidopsis thalia
Soybean raffinose
Corn raffinose syn
Amino acid sequenc
Arabidopsis thalia
Arabidopsis thalia
Soybean raffinose
Arabidopsis thalia
Arabidopsis thalia
Novel human diagno
Novel human diagno
Novel human diagno
Pit-1extraRSTM/cy
Lettuce resistance
Alpha 2-Macroglobu
Human alpha-2-MK.
Human alpha-2-MK.
Human LDL receptor
Drosophila melanog
Lettuce resistance
Chlamydia pneumoni
Sequence encoded b
Truncated PLT SVGS

CC gene can be used for expression in a plant for the production of
CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence represents raffinose
CC synthase from soybean.

XX Sequence 750 AA;

Query Match 100.0%; Score 3996; DB 20; Length 750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTPVPRKISVNDGKLVHGKTLTGVPNDVNLTPGSGRGIVTGAFGVGTATASHSKSLHVEPM 60
DB 1 mtpvprkisvndgklvngkltlvgpndvnltpgsgrgivtgaavgatashskslhvepm 60
OY 61 GVLGSLRFMCCEFRFKLMMWMTORMGTGCRDVPLETOFMILSKSESETDGENSPIIYTVLP 120
DB 61 gvlgsrlfmccefrfkllmmwmtormgtcgrdvpлетofmilsksestdgenspiiytlp 120
OY 121 LLEGQFRAVLQGNDRKNEIETCLESQDNAAVETDQGLHMYVMHAGNPREVINQAVKAEKH 180
DB 121 llegqfraelqgnrkneietclesqdnaveetdqlhmyvmhagnprevinqavkavekh 180
OY 181 MOTFLHREKRRKRLPSCLDFGMCWTDAFYTDVTAEGVEEGLKSLSQGGTPPREFLIIDGMQ 240
DB 181 motflhrekrkrlpscldfgmcwtdafytdvtaegveeglkslsqggtpprefliiddgmq 240
OY 241 OIENKAKDATECLVOEQAOFATRLTGIEKNTKFKKLLQNNQNSGLKHLVHGAKQHNHV 300
DB 241 oienkakdateclvoeqaofatrltgiekntkfkklqnneqmsglkhlvngakqhnvk 300
OY 301 NVYVNHALAGYWGKVPRAATGMEHVDTLAYPVOSPGVLGNQPRIVMDSLAVHGLGLVHP 360
DB 301 nvvvnhalagygwkvpaaatgmevdtalayvospgvlgngprivmdslavhglglvhp 360
OY 361 KVFENYNEHLAYLASCGVDGKVDVQNIETLGAGHGRVSLTASYHHALEASIASNFT 420
DB 361 kvfenynelhaylasccgvdkvkvqnietlgaghgrvsltasyhhalessiasnft 420
OY 421 DNGCIACMCHNTDGLYSKQTAIVASDDFPRPASHTIHISVAVYSLTGEMQPDW 480
DB 421 dngciacmchntdglysakotaivasddfprpashtihisvavysltgelmqpdw 480
OY 481 DMFHSHPAADYHAARAIIGCCPIYVSKPGMHNFDLKTIVLPDGSYLRAQLPGRPTRD 540
DB 481 dmfhsihpaadyhaaraiaigccpiyvsdkpgmhnfdllktilvpdgsylraqlpgrrprd 540
OY 541 SLFVDPARDRTSLKILWNLKNCSGVGVFNCQAGACWCKIEKKRIHDTSPGTLTASVCS 600
DB 541 slfvdpardrtsllkilmnlkncsgvvgvfncqagacwkiektrihdtspgtltasvas 600
OY 601 DVDLITOVAGAMLDITVYARSGEVRRLPKGVSIPTLVKLEPRLHPCIOBIASI 660
DB 601 dvdlitovagamllditvayarsgevrlpkgsviptrlvkleprlhpciobiasi 660
OY 661 SFAAIGLDMFTMGAVEQVEIHNRAATKTIALSVGRGRFQVSSOBRLKCVGGAETD 720
DB 661 sfaaigldmftmgaveqveihnraaktialsvgrgrfgyvssqbrlkcvvggaetd 720
OY 721 FNTDSETGLTFSIPVSPEDMTKMSLEIOY 750
DB 721 fntdsetgltfstipvspeemtkmsleioy 750

RESULT 2

AAT70980 AAY70980 standard; Protein; 751 AA.

XX AAY70980;

XX 09-AUG-2000 (first entry)

DE Wheat raffinose synthase from clone wlm24.pk0021.h1.

XX Wheat; raffinose synthase; raffinose saccharide; soybean;
KM clone wlm24.pk0021.h1; nutritional; soy protein.

XX Triticum aestivum.

PN WO200024915-A2.

PD 04-MAY-2000.

PF 22-OCT-1999; 99WO-US24923.

PR 23-OCT-1998; 98US-0105451.

PA (DUPO) DU POINT DE MEMOIRS & CO E. I.

PI Allen SM, Hltz WD;

XX WPI; 2000-350754/30.

DR N-PSDB; AAD00337.

PT Nucleic acids and encoded proteins involved in the biosynthesis of
PT raffinose, useful for producing soybean seeds with a reduced raffinose
PT content and therefore improved nutritional quality -

PS Claim 22; Page 52-54; 58pp; English.

CC The present sequence is a raffinose synthase from
CC clone wlm24.pk0021.h1 isolated from a wheat seedlings cDNA
CC library wlm24. Raffinose synthase is involved in the biosynthesis
CC of raffinose and higher homologues in the raffinose saccharide family
CC from sucrose. The present sequence is useful for reducing the raffinose
CC saccharide content of soybean seeds which improves the nutritional
CC quality of the soy protein products derived from them.

XX Sequence 751 AA;

Query Match 74.6%; Score 2979.5; DB 21; Length 751;
Best Local Similarity 72.1%; Pred. No. 2.8e-300;
Matches 544; Conservative 94; Mismatches 109; Indels 7; Gaps 5;

OY 1 MTPVPRKISVNDGKLVHGKTLTGVPNDVNLTPGSGRGIVTGAFGVGTATASHSKSLHVEPM 60
DB 1 mtpvprkisvndgklvngkltlvgpndvnltpgsgrgivtgaavgatashskslhvepm 60
OY 61 GVLGSLRFMCCEFRFKLMMWMTORMGTGCRDVPLETOFMILSKSESETDGENSPIIYTVL 118
DB 61 gvlgsrlfmccefrfkllmmwmtormgtcgrdvpлетofmilsksestdgenspiiytlv 118
OY 119 LPLEGQFRAVLQGNDRKNEIETCLESQDNAAVETDQGLHMYVMHAGNPREVINQAVKAE 178
DB 119 lplegqfraelqgnrkneietclesqdnaveetdqlhmyvmhagnprevinqavkave 178
OY 179 KHMOTFLHREKRRKRLPSCLDFGMCWTDAFYTDVTAEGVEEGLKSLSQGGTPPREFLI 238
DB 179 khmotflhrekrkrlpscldfgmcwtdafytdvtaegveeglkslsqggtpprefli 238
OY 239 WOQIENKAKDATECLVOEQAOFATRLTGIEKNTKFKKLLQNNQNSGLKHLVHGAKQHN 298
DB 239 woqienkakdateclvoeqaofatrltgiekntkfkklqnneqmsglkhlvngakqhn 298
OY 299 VKNVYVNHALAGYWGKVPRAATGMEHVDTLAYPVOSPGVLGNQPRIVMDSLAVHGLGL 358
DB 299 vknvyvnhalagygwkvpaaatgmevdtalayvospgvlgngprivmdslavhglgl 358
OY 359 HPKVFENYNEHLAYLASCGVDGKVDVQNIETLGAGHGRVSLTASYHHALEASIASN 418
DB 359 hpkvfenynelhaylasccgvdkvkvqnietlgaghgrvsltasyhhalessiasn 418
OY 419 FTDNGCIACMCHNTDGLYSKQTAIVASDDFPRPASHTIHISVAVYSLTGEMQPDW 478
DB 419 ftdngciacmchntdglysakotaivasddfprpashtihisvavysltgelmqpdw 478

Db 418 fdngsciscmhntdmlysakqtavrasddfyrpdrpashvhisvayntclfigefmq 477
QY 479 DDMFMSLHPADYHAARAIGCCPIYVSDKGNHNPOLLKVLVPGSVLRADLPGRPT 538
Db 478 dwwmfmslhpadaeyhaaraigccpiyvsdkpgnhnpollkvlvpdgsvlraqlpgprpt 537
QY 539 RSLFVPADPDRTSLLKINLNKSCGVVGVFCOGAGMCKIEKTRHDTSPGTLTASVC 598
Db 538 rdclfsdparpdgasllkikvnmhkacagvvgvfcogagwcrvkktrihdeapgtltsvvr 597
QY 599 ASDVDLITQVAGG-EMLGDTIYVAYRSGEVIRLPKGVSIPTVLKLEELHFHFCPI 657
Db 598 aevvegltaetqddctgdaavvvtlnagelivprgatlvplkileyelthvcpvava 657
QY 658 PSLISFAIGLDMFTNGAVGEQVEIHNRAATK-TIALSVGRGRGVYSOSRPLKCVYG 716
Db 658 pdlfsfipglilhmfnagaveecvrtneaddaavalrvgrgtrigaycstrpraksids 717
QY 717 AETDFNIDSETGLTFISIPVSPBEMTMSIEIQV 750
Db 718 advefigydatgtlvtdvpvpeemrywcllelrv 751

RESULT 3

AA70975 standard; protein; 756 AA.

AA70975;

09-AUG-2000 (first entry)

Corn raffinose synthase #2.

Corn: raffinose synthase; raffinose saccharide; soybean; nutritional; soy protein.

Zea mays.

MO200024915-A2.

04-MAY-2000.

22-OCT-1999; 99WO-US24923.

23-OCT-1998; 98US-0105451.

(DUPO) DU PONT DE NEMOURS & CO E. I.

Allen SM, Hitz WD;

WPI: 2000-350754/30.

N-PSDB: AAD00332.

Nucleic acids and encoded proteins involved in the biosynthesis of raffinose, useful for producing soybean seeds with a reduced raffinose content and therefore improved nutritional quality -

Claim 2; Page 36-39; 58pp; English.

The present sequence is a raffinose synthase from a contig of clones cbn10.pk0034.e8 and cbpc24.pk0003.h7 isolated from corn cDNA libraries cbn10 and cbpc24 respectively.

Raffinose synthase is involved in the biosynthesis of raffinose and higher homologues in the raffinose saccharide family from sucrose. The present sequence is useful for reducing the raffinose saccharide content of soybean seeds which improves the nutritional quality of the soy protein products derived from them.

Sequence 756 AA;

Query Match 71.9%; Score 2874; DB 21; Length 756;
Best Local Similarity 68.8%; Pred. No. 2.8e-289;

Matches 521; Conservative 102; Mismatches 126; Indels 8; Gaps 4;

QY 1 MMYTPKISVNDCKLIVHGGTILITGPDNVVLPFGSGRLVYGAVVGNATSKSLHFFM 60
Db 1 mlytpkivsdgrrlvtgrgtvltvpcdhvsaahagagldvgatvgalagaekshvltf 60
QY 61 GVLESLRFMCCRFKLMWMTQRMCTGHDVPLETQFMILSKESKESDTDENSPIITYLLP 120
Db 61 gvlrlrdlfcilfrfkvwmtqrmgysgrclvretqfmilvcpasdgddadp-ayvmlp 119
QY 121 LLEGGFRALVQGNDRKNEIEICLESQDNVETDQGLHMYMAGTNPPEVINOAKAYEKH 180
Db 120 llegqfiraalqndetdeiqiciesgdkaqctgaahmylhagdnrpfvtaavkavekh 179
QY 181 MDTFLREKRLPSCIDMFMGCTMDAFYTDVYAEVEBGLKSLSGGTPRFLITDGMQ 240
Db 180 lqtfhnrdrkklksrlfdwfgwctwdaftydvlaadvknqslsksggpprflitddgwg 239
QY 241 QIENKAKDAITECLVDBGAOFATRLTGIKENTKFOKKLONN----EOMSGLHVLHGAKH 296
Db 240 qlasenkpdpnvavgeaqfaasrltgikentkfqkdpdgdgdegaagylkrlvaekda 299
QY 297 HMYKNYVNHALAGYGVGP-AATGMEHPTALAYPVOSPGVLGNCPDIYMDSLAYHGL 355
Db 300 hgykyvwhamaagywgytprtaaglamelyepalypryqspgvngnqpdlymdslavgl 359
QY 356 GLVHPKKVFNFEYNEHAYLASGVDCVKDVONITETLAGCHGVSLTRSYHHALEASI 415
Db 360 glvhprrvrdfygelhaylasgvdvkvqnllletlgaqhgsvvaltrayhraleasv 419
QY 416 ASNFDTNGCIACGCHNTDGLYSAKOTAIYRASDDYFPDPASHTHISVAYNSLFGEEF 475
Db 420 arsfpndgscmhnsdmlysaarqtavrasddfyrpdrpashvhisvayntclfigef 479
QY 476 MOPDMFMSLHPADYHAARAIGCCPIYVSDKGNHNPOLLKVLVPGSVLRADLP 535
Db 480 mqpdmfmslhpadaeyhaaraigccpiyvsdkpgnhnpollkvlvpdgsvlraqlpg 539
QY 536 RPTRDSLFPVDPADPDRTSLLKINLNKSCGVVGVFCOGAGMCKIEKTRHDTSPGTLTASVC 598
Db 540 rptdrclfsdparpdgasllkikvnmhkacagvvgvfcogagwcrvkktrihdeapgtltsvvr 597
QY 596 SVGASVDLITQVAGG-EMLGDTIYVAYRSGEVIRLPKGVSIPTVLKLEELHFHFCPI 653
Db 600 svradvdalariagdgqgdgetvvyahrtrelrvlprgvalpvlqplqyevfhvcp 659
QY 654 QETAPGISFAIGLDMFTNGAVGEQVEIHNRAATK-TIALSVGRGRGVYSOSRPLKCV 713
Db 660 ravvpavstapvylldmfnagaveecdvlsdvgkamalrvgrgtrigaycstrpractl 719
QY 714 VGGAEITDFNIDSETGLTFISIPVSPBEMTMSIEIQV 750
Db 720 ldsaevefsydygtlvsdvltvpegeylvclclmi 756

RESULT 4

AA70977 standard; protein; 763 AA.

AA70977;

09-AUG-2000 (first entry)

Rice raffinose synthase from clone rls72.pk0020.d0.

Rice: raffinose synthase; raffinose saccharide; soybean; clone rls72.pk0020.d0; nutritional; soy protein.

Oryza sativa.

MO200024915-A2.
04-MAY-2000.

Query Match	70.8%;	Score 2827.5;	DB 21;	Length 763;
Best Local Similarity	68.2%;	Pred. No. 1.9e-284;		
Matches 523;	Conservative 95;	Mismatches 128;	Indels 21;	Gaps 7

```

Oy      592  TIASVASCPADITLQVAGAE---WLDPTIVYVRSESVIRLPRKGSIVPTLKATVEELF 648
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      600  tlgavrvaddvalaqvagvgdgvgdgavvyahararelrvlprigpaaplrvltglaleyevf 659
Oy      649  HRPDIEI-----PSISPAIIGLIDMPNMGSAVEOYEIHNNAARKTILASVRKSGPFGV 703
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      660  hvcevtalaaapgaavatalapqyllidmfngagvceavdaaaa---valrvygcgffga 716
Oy      704  YSSORPLKVCYGGAETDFRNDYSEGLTTFSPVSPBEMTRMSLIEIYV 750
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      717  yfstrparcaldaadvgltygdvdlvavdlpvpeqemuyfwehelnv 763

```

Result	5	
AA70976	5	
ID	AA70976	standard; Protein; 770 AA.
XX		
XX	AA70976;	
XX		
XX	09-AUG-2000	(first entry)
XX		
XX	Rice raffinose synthase from clone rls24.pk0017.g10.	
DE		
XX		
XX	Rice; raffinose synthase; raffinose saccharide; soybean;	
KM		
KM	clone rls24.pk0017.g10; nutritional; soy protein.	
XX		
OS	Oryza sativa.	
XX		
XX	Key	Location/Qualifiers
XX	Misc-difference 100..101	
XX		/note= "Encoded by a region of 105 nucleotides
XX		represented as base pairs 693-797 in AAD00333"
XX	Misc-difference 101..110	
XX		/label= Unknown
XX		/note= "Encoded by NNNNNNNNNNNNNNNNNNNNNNNNNNNNN"
XX	W0200024915-A2.	
XX		
XX	04-MAY-2000.	
XX		
XX	22-OCT-1999;	99WD-US24923.
XX		
XX	23-OCT-1998;	98US-0105451.
XX		
XX	(DUPO) DU PONT DE NEMOURS & CO E I.	
XX		
XX	Allen SM, Hitz MD;	
XX		
XX	WP1: 2000-350754/30.	
XX	N-PSDB; AAD00333.	
XX		
XX	Nucleic acids and encoded proteins involved in the biosynthesis of	
XX	raffinose, useful for producing soybean seeds with a reduced raffinose	
XX	content and therefore improved nutritional quality -	
XX		
XX	Claim 2; Page 40-42; 58pp; English.	
XX		
XX	The present sequence is a raffinose synthase from	
XX	clone rls24.pk0017.g10 isolated from a rice infected leaf cDNA	
XX	library rls24. Raffinose synthase is involved in the biosynthesis	
XX	of raffinose and higher homologues in the raffinose saccharide family	
XX	on sucrose. The present sequence is useful for reducing the raffinose	
XX	saccharide content of soybean seeds which improves the nutritional	
XX	quality of the soy protein products derived from them.	
XX		
XX	Sequence 770 AA;	
XX		
XX	Query Match	53.0%; Score 2117.5; DB 21; length 770;
XX	Best Local Similarity	52.8%; Pred. No. 1.7e-210;
XX	Matches 408; Conservative 111; Mismatches 210; Indels 43; Gaps	
XX	1 MTVTPKISVNDGFLVYHGKTTILTGVPDNYVLTGSGRGLVTGAIVGATASHSKSLHVEPM 60	

```

Db      1 mtyssvkvaggselvtgrtvlsgvpeavrrasaagapvdgflggdlaepssrhwsl 60
QY      61 GVEGLEFMCCFRFKLMMTQRMGTCGRDVPLETOFMLIESKESSETDGENSPITTYVLLP 120
Db      61 gawrgmifmccfllkwwagmqmgekgyddphtqfllveaxxxxxx-----xxxxlvlp 114
QY      121 -LLEGOFRVAVLOGNDK--NEIEICLESQD--NAVETOGLHMVYMAHGTNPEEYINAA 173
Db      115 pasesgafrpslsggaggaadelqlcvesgdagtraaftalvt---gpdsdpfaalaga 172
QY      174 VKAVEKHMOTFLHREKKRLPSCLDWFGMCTWDAFYTDVAEGVEEGLKSLSOGTPRRL 233
Db      173 vaaakscilffiraeakklpjglvdgfwctwdafygdvtrgeqveaglsrltsggappkvt 232
QY      234 IIDDGMOOI--ENKADATECLVQEGAOFATRLTGIKENTKROKLIQNNQMSGKHLVH 291
Db      233 liddgwsvgtldhnpddtgaadkdkpjlartltgikenskfg---dgdppaagiktvtvr 289
QY      292 GAKOHNVKNVYVWMAHLAGYWGVRKPAATGMEHYDPAALAYPVQSPGLGNODIYVDSLA 351
Db      290 aakekyglkyvyywhaitgywgvvtrpgvamegyhsnmqfnpvpsgvneepgmktldvtl 349
QY      352 VHGGLVHREKKVFNFYNEHLHAYLASCGVDGVKVDVQNIETLQAGHGRVSLTRSYHNL 411
Db      350 tgglgivhpravyrfydelhaylaagvdygkvvdvqclletlgaqgyrvtlrrqfhgal 409
QY      412 EASIAFNFDNGCIACMCNHTDGLYSAKOTAIYRASDDPRPDPAHSHTHISSVAYNSLF 471
Db      410 dasiafnfengliacmshntdalycaakqavrasddtprdpshthlssavaynsyf 469
QY      472 LGEMQPDMDMFHSLHAPADYHAARATIGCPITYVSDKPGNNFDLLKKLVLPDGSVLA 531
Db      470 lgefmpdmdmfhslhpadgyhsaralsgpyvdsdapgkhnfelkkmvlpdgsvlra 529
QY      532 QLPGRTRBSLVDPARDRTSLTKTMINLKSGGVGVFNCGAGMCKTEKTRIRHDTSPG 591
Db      530 wlpgrtrbslvdpardrtsltkminlksggvlgvynrcqgaavssvekknlfhktgae 589
QY      592 TLTAASVCAADVDLITQVA--GAEMLGDTIYAYRSGEIVRLPKGVASIPVTLKYLEFELFH 650
Db      590 alsqykgsgdvhladaatdsemdcavyrtnasadiwlpnagaalpslkyvlendhlvt 649
QY      651 CPTQETAPSISPAIGLDMETGAVEQVEIHNRAAKT----- 690
Db      650 spikdlapgftrfapjglvdmfinsgaavegltyhridyvkylsngsaastlpelqslsqa 709
QY      691 -IALSVGRGRGREGVSSORPLKCVVGAENDFNDSIEGLTTSIPVSPES 740
Db      710 gltvmevtrgcykfagysvtrpikcmllgsaqvefitydsassgvlldlempke 761

```

RESULT 6
AAV70981
ID AAV70981 standard; Protein: 841 AA.

AAV70981;

09-AUG-2000 (first entry)

Wheat raffinose synthase from clone wlm96.pk033.h5.

Wheat, raffinose synthase; raffinose saccharide; soybean;
clone wlm96.pk033.h5; nutritional; soy protein.

Triticum aestivum.

WO200024915-A2.

04-MAY-2000.

22-OCT-1999; 99WO-US24923.

XX

```

PR      23-OCT-1998; 98US-0105451.
PA      (DUPLO ) DU PONT DE NEMOURS & CO E.I.
XX      Allen SM, Hiltz WD;
PI      WPI: 2000-350754/30.
XX      DR N-PSDB; AAD00338.
XX      N-PSDB; AAD00338.
XX      Claim 2; Page 55-58; 58pp; English.
CC      The present sequence is a raffinose synthase from
CC      clone wlm96.pk033.h5 isolated from a wheat seedlings cDNA
CC      library wlm96. Raffinose synthase is involved in the biosynthesis
CC      of raffinose and higher homologues in the raffinose saccharide family
CC      from sucrose. The present sequence is useful for reducing the raffinose
CC      saccharide content of soybean seeds which improves the nutritional
CC      quality of the soy protein products derived from them.
XX      Sequence 841 AA;
SQ

```

Query Match 52.88; Score 2110.5; DB 21; Length 841;
Best Local Similarity 53.18; Pred. No. 1.1e-209;
Matches 407; Conservative 123; Mismatches 213; Indels 23; Gaps 10;

```

QY      1 MTVPRKISVNDGKLVYHAKKTITLGGVDPNVVLLPSSGGLYVGARVATASHSKSLHFFPM 60
Db      81 mtiesvsvlsgelstvgtrvlsgvpdavsaspaarqpvddvlgadlqgparthvsl 140
QY      61 GVEGLEFMCCFRFKLMMTQRMGTCGRDVPLETOFMLIESKESSETDGENSPITTYVLLP 120
Db      141 ghmgvtrfmacfllkwwagmqmgekgyddphtqfllveaxxxxxx-----xxxxlvlp 114
QY      121 LLEGOFRVAVLOGNDK--NEIEICLESQD--NAVETOGLHMVYMAHGTNPEEYINAA 173
Db      199 lvegafraslsggaggaadelqlcvesgdagtraaftalvt---gaadsdpfaalaga 256
QY      176 AVEKHMOTFLHREKKRLPSCLDWFGMCTWDAFYTDVAEGVEEGLKSLSOGTPRRL 235
Db      257 avrscldgtrfpraeakkrlpaldvfygwtwdafygdvtrgeqveaglsrltsggappkvt 316
QY      236 DDCGMOOI--ENKADATECLVQEGAOFATRLTGIKENTKROKLIQNNQMSGKHLVHGA 293
Db      317 ddgqsvgtldkqspdlasagaagkspplprltgikenskfgsg--dpatatgletlvraa 375
QY      294 KOHNVKNVYVWMAHLAGYWGVRKPAATGMEHYDPAALAYPVQSPGLGNODIYVDSLA 351
Db      376 kekylkyvyywhaitgywgvvtrpgvameaayrsmqfnpvpsgvneepgmktldvtl 435
QY      354 GLGLVHREKKVFNFYNEHLHAYLASCGVDGVKVDVQNIETLQAGHGRVSLTRSYHNL 411
Db      436 glglvhpqavhrfydelhaylaagvdygkvvdvqclletlgaqgyrvtlrrqfhgal 495
QY      414 SIASNFNDNGCIACMCNHTDGLYSAKOTAIYRASDDPRPDPAHSHTHISSVAYNSLF 471
Db      496 svaknfdngliacmshntdalycskqlavrasddtfrcaavshthlaaavaynsf 555
QY      474 EFMOQPDMDMFHSLHAPADYHAARATIGCPITYVSDKPGNNFDLLKKLVLPDGSVLA 531
Db      556 efmoqpdmdmfhslhpadgyhsaralsgpyvdsdapgkhnfelkkmvlpdgtvltarl 615
QY      534 PGRPTRBSLVDPARDRTSLTKTMINLKSGGVGVFNCGAGMCKTEKTRIRH-DTSPGT 592
Db      616 pgrptrbslvdpardrtsltkminlksggvlgvynrcqgaavssvekknlfhktgae 675
QY      593 LTAASVCAADVDLITQVA--GAEMLGDTIYAYRSGEIVRLPKGVASIPVTLKYLEFELFH 649
Db      676 ltgavtrgdvhladaatdsgagsgdcavyrthgagdlvlpdgvalpslkyvlendvlt 735

```


Best Local Similarity 39.7%; Pred. No. 5,6e-139;
Matches 302; Conservative 127; Mismatches 262; Indels 69; Gaps 15;

```

QY 8 SYNDGKLVHGKTLITLGPVNVLTLPSSGRLVTGAFVGTASHSKSLHVPKGLR 67
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 22 slcnstlkvngqvllsgpknvlltpctydtltgctflgfhatspkrharpjgqlknls 81
QY 68 FMCFPRKLMWMTQRMGTGRDVPLETOFMLIESKESETDGENSPILYTVTLTLLESGQR 127
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 82 ftsifrkvwttltwtsnrdtletetqfmlmishp-----yvlfrlpilqptr 130
QY 128 AVIAGNDKNEIETCLESQDNAAVETDGLHVMYHAGTNPEVINQAVKAEKMQTEFLHR 187
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 131 aslqpheddvavcvesshvlassfdvtvylhagdnptlvekamrvrvalhgsfkll 190
QY 188 EKKRLPSCLMFGMCTMDAFTYDTYTAGVEGLKSLSGQTPRRLTIDGQOIKENAK 247
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 191 eeklvpmvdkfjwctclafyltvhpqyregvklvdggcpqfvllddgwqclshdsd 250
QY 248 DATECLVQ--EGAQFATRLTGIKENTFKOKLONNEOMSLGKTLHVGAQOH--NVKNVY 304
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 251 pekqgmqtvgqgmprcllsyeenykf-----tsyekgklkfvelkeefsgvseyvyy 306
QY 305 WHALAGVGVKPAATGMEHYDTALAVPVQSPGLGNQDIWDSLAVHCLGVHPRKYV 364
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 307 whalcgwgvgrpvagm--eaaavekprlleglkgmedlavdklvnngvgyvvpelvy 364
QY 365 NFNELHAYLASGVGVKVDYQNIETLGAHGGRVSLTRSHALHLSIASNFIDNGC 424
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 365 emeyghahlesagldgvkvdvllhlemcekyggrvdmaakaykaltasvthkfyngyn 424
QY 425 IACMCHNTD--GLYSAKOTATVRSDDFVRPDP-----ASHTHISSVAVNSLPL 472
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 425 lasmebondfmlilgtelaistgrvgddfwctdpygdngfclwqclmnhv---caynsllm 481
QY 473 GERMODPMWFHSLHRAADYHAAARIGCPTIVSKPGNHNFDLIKTLVDPGSLVRAQ 532
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 482 gnlihpwdmftgltpcaatlaasraistgprlyistdvgnhnfellktlalpdsjllrce 541
QY 533 LPRGPTRDSLFPVARDRTSLKTIWNLKSGVGVFNGOGAWCKIEKTRHDSPT 592
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 542 hyalprtdcfadplndgkmtlkiwnlkytgylvfncqgggwtf-----e 588
QY 593 LNASVCASD---VDLITVAGAEW-----LGDITVAVRSGE---VIRLPKGVSI 636
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 589 itsnkcaaeishvstklklkldewdsgknplsliegvtlfaayfsgakklillsapsdde 648
QY 637 PVTLKYLELELHFHFCIOEI-APSISFAALGLDMFNTGCAVEQVEIHNRRAKTITALSV 695
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 649 ealslepfntellitvsvptvlpqksvktapldglvmmlntgavqslafde--gqnllvevyl 706
QY 696 RGRGRGVYSSQRLPKCVGAGTDPNYSSETGLTTFESIP 735
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 707 rgtgemrvyasekptrctldgkevdlcy--egsmwnlqyvr 744

```

RESULT 9
AA32074
ID AA32074 standard; Protein: 777 AA.
XX
AC AA32074;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mustard raffinose synthase.
XX
KM Raffinose synthase; mustard; transgenic plant.
XX
OS Brassica juncea.
XX
FH Key Location/Qualifiers
FT Misc-difference 210 /note= "encoded by ACR"
FT

```

XX EP953643-A2.
PN
XX
XX 03-NOV-1999.
PD
XX
XX 27-APR-1999; 99EP-0107430.
PF
XX
XX 30-APR-1998; 98JP-0120550.
PR 30-APR-1998; 98JP-0120551.
PR 04-DEC-1998; 98JP-0345590.
PR 10-DEC-1998; 98JP-0351246.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
PA
XX
XX Watanabe E, Oeda K;
PI
XX WPI; 1999-593144/51.
DR N-PSDB; AA220209.
XX
XX New sense and antisense genes, useful for altering the level of
PT raffinose in food plants -
PS
XX
XX Claim 26; Page 29-31; 55pp; English.

```

This sequence represents mustard raffinose synthase, a protein that can bind a D-galactosyl group through an alpha(1-6) bond to the hydroxy group attached to the carbon atom at the 6-position of the D-glucose residue in a sucrose molecule to form raffinose. cDNA (see AA220209) encoding the enzyme was isolated from mustard (Brassica juncea) leaf cDNA by PCR. Probes or primers generated from plant raffinose synthase genes (see AA220207-10) may be used to obtain other raffinose synthase genes by labeled detection or amplification (claimed). These genes may be used to control the levels of raffinose produced in plants. Antisense genes can be used to knock out existing gene activity, and sense genes to increase the level of gene activity. The resulting transgenic plants may be used as a food source to alter the growing conditions for gut enterobacteria, providing general health advantages.

Sequence 777 AA;

Query Match 35.4%; Score 1415.5; DB 20; Length 777;
Best Local Similarity 38.2%; Pred. No. 2.1e-137;
Matches 296; Conservative 133; Mismatches 279; Indels 67; Gaps 17;

```

QY 5 KRISVNDGKLVVHGKTLITGVPDNNVLT-----PSSGRGLVTGAFVGTAA--SHSKSL 55
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 22 plfrlegsdllanghvvltvdpvntvltaspyladdgdepvdasagsfifnldgeprsr 81
QY 56 HVPKGVLEGLRPMCCFRFKLMWMTQRMGTGRDVPLETOFMLIESKESETDGENSPILY 115
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 82 hvastlgklrtdrlfmslfrfkvwttlwvsgksdlenetqlillen-----sgsgrp--y 134
QY 116 TVLLPLEEGOPRAVLQSGNDKNEIETCLESQDNAAVETDGLHVMYHAGTNPEVINQAVK 175
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 135 vlllpilllegsfrrsfqgedddavavcvesgvtgsefgrvyvvnhaagddpriklvkdamk 194
QY 176 AVEKNMQTEFLHREKKRLPSCLDWFGMCTMDAFTYDTYTAGVEGLKSLSGQTPRRLII 235
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 195 vrvvhnmtfklleektpgvlvktgwtwdaftltnpddghvkvkclvdggcpqfvlvl 254
QY 236 DDGMOQIENKAK-----DATECLVQEGAQFATRLTGIKENTFKOKLONNEOMSG--GLKHLV 290
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 255 ddgwsisghdsdgidvqgmactv-ageqmporllkfgentkftdyvsprkdkneymkafv 313
QY 291 HGAKOHHN--VKNVYVHVALAGVGVKPAATGMEHYDTALAVPVQSPGLGNQDIWDSL 349
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 314 rdlkeefstvdylvvnhalogvysgqlrpgaprlp--pstlvrepelspgllmqdlaavdk 371
QY 350 LAVHGLGLVHPRKYVFNFYNELHAYLASGVGVKVDYQNIETLGAHGGRVSLTRSHYH 409
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 372 lvdctgigfvsdpmanefyegjshnlgnvgldgvkvdvllhlemcekyggrvdmlakayfk 431

```

Oy	410	ALESAISNFENDNCIACOMCHNDGVL-SAKORAIYASDFPDRPAS-----	457
Dd	432	altssvknfhfgangvisamelmclndfmlliglealslgvgvdffmctbdpsglingtlylqgc	491
Oy	458	HTHHSVAVNSLFLGEPMOPDMDHSLHPADYHAAPAAIGCPIYVSDFKGNHNFDL	517
Dd	492	hmhv--caynsilmwgmflfqdwmtfstbtpcaeffaasarsls9p9rlylsdcvgvgnld	548
Oy	518	LKRLVLPDGSVYLRAOLCGRPFRDLSFPDPARDRSLILKNLNKSGSIVGVFVMQGAWC	577
Dd	549	lkrlvlpgdsjlrcchyalprdtlrfeodphldgkltmklnmlnkylgl:afagncggswc	608
Oy	578	KIEKKTRIHDTSPGTLTASVASVDLTVOYAAGW-----LGDTIYYAARSGE	626
Dd	609	retrrngvfsgcvntlatatnkdav-----ewmsgmplsvenveefalfisstk	659
Oy	637	VIRLRGVSIVPTLVLEFEELHFHCPIQIEL-APSISFAALGLDMFTMGVAVEQLHNHR	685
Dd	660	lvlsgrddlelttlepfkfelltsvspvtlyegssvqlapjglylmnltsгалsrlsyhe-	718
Oy	686	AATKIHALSVGRGRFGFYSSQRPLKCVCYGGAETDFMYNDSTGTLTTSIYVSPEE	740
Dd	719	--esvelgvrgdaetivyaarskascldgeveley--eeswmvyqpvsape	768

RESULT 10
 AAM53570 standard; Protein; 784 AA.
 ID AAM53570 standard; Protein; 784 AA.
 XX
 AC AAM53570;
 XX
 DT 06-JUL-1998 (first entry)
 XX
 DE Cucumber raffinose synthase.
 XX
 KW Cucumber; raffinose synthase; sucrose; galactinol.
 XX
 OS Cucumis sativus.
 XX
 PN JPI0084973-A.
 XX
 PD 07-APR-1998.
 XX
 PF 28-APR-1997; 97JP-0111124.
 XX
 PR 26-JUL-1996; 96JP-0198079.
 PR 26-APR-1996; 96JP-0107682.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 DR MPI; 1998-264858/24.
 DR N-PSDB; AAM22250.
 XX
 PT Raffinose synthase gene - useful for preparation of raffinose in
 PT transformed plant
 XX
 PS Claim 3; Pages 17-20; 26pp; Japanese.
 XX
 CC The present sequence is cucumber raffinose synthase, which
 CC forms raffinose from sucrose and galactinol, has an optimum pH of
 CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
 CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
 CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
 CC iodoacetamide, N-ethylmaleimide and myoinositol.
 XX
 SQ Sequence 784 AA;

Query Match	34.98;	Score 1394;	DB 19;	Length 784;
Best Local Similarity	37.68;	Pred. No. 3.7e-135;		
Matches 289;	Conservative 143;	Mismatches 285;	Indels 52;	Gaps 15;

```

0Y 8 SVN0GKLVNHEKTTLLTEGVDPNNVLTLPSSGGL-----YNGAFNAAASNSKSLHNEPMV 62
Db 26 a1dgsdfvlnghsflsdvpenlvaaspryslksdpvsyvcfvgffaaesdrfnhvsyjk 85
0Y 63 LEGRFMCCEFRFKLMMWTORHGTGCRBPVLETOFMLESKESETDGENSPITVTLPL 122
Db 86 lkdfrmslfrfkwtvethwvgrngslsestqvlleked-----gpr--yvfllprlv 138
0Y 123 EGFGRVALGNDKNEIEICLESQDNAVETDQGLHMYTMAAGTRPEVYNDAKVAEKHQ 182
Db 139 egrftrsqpdddfvdvcesgsskvvdasfrsmlylhagddprfalvkeamkivtrthlg 198
0Y 183 TFLNREKRRLPSCLDNMGKMTWIAFEYDVAEAGEESKLSGGRPRRLTIDDSMOOI 242
Db 199 ftllleektrpgrvdklqwcwdaftytltrpgrvlegrvnlvdgscrprvliiddgwsj 258
0Y 243 ENKADATECJLVQ-----GAOFATRLTKIKENTKFO-----KKLONNDSMGLKHLNGAK 294
Db 259 ghdsprlckegmqvtvgcegmprcllkfgenykrfdryvnrpketrpagaqgmakafideljk 318
0Y 295 -QHNNKVVNYMHALAGWGSVKPRATGEMEHYDTALAYPQSGEVLGNORDIYMSLANH 353
Db 319 gefkvevhuywhalacgywglrqpgrlp--earvlgpralspqlgmthedaivdklvh 376
0Y 354 GGLVNHPRKKVNFNFNEHLNATYLAASGVDSYKVDYONIIETLAGHNGGVSLTSSYHNAL 413
Db 377 kvglyrpkraeemegqhabhekvgldrvkdvhlhlemiceodgyrvdlaakayukamtk 436
0Y 414 STASNFOTNGCIACGCHNTDGLY-SAKOTAIVRASDDEFPDRPAS-----HTIH 461
Db 437 slnkhfrkngvylaamehondlmfllgtaeislgtygddftwctdrpsgrngftwlgscimhw 496
0Y 462 ISSVAYNSLFTGEFMQPDMDFNHSLHPRADYHAAARAGCCPIYVSXKPCNNHNFDLKL 521
Db 497 -----candslmngfnfhpdmfmgsthrpcaafhaasraizsgrlyvdsavgkhmfidlkli 553
0Y 522 VLPDGSVLRADLPGRFPTDLSFLVDPARNRISLKIWNLMKSGGVGVGVEFCOGAGWKIEK 581
Db 554 vlppdgslltrseyalprtdclfedrllhngelmklwlnkftvgylagfncqggwccetr 613
0Y 582 KTRINDSPGTLTASVCASVDVLT--OVAGAEMVLGPTIYAVYAGSGVIRLPRKGVSIPTV 639
Db 614 tnqctsgykrvstxtnpkdlemhsgenrlsegyvktfalyakklkllskpsgdldla 673
0Y 640 LKVLFEFLFHFNCPIQD-IAPSISEAIGLDMFNTGCAVEOYELHNRAAKITLALSVGR 698
Db 674 ldrfefeiltsvprklkltqtslhfabrjlymnltsaigavdvddds--sveilgykc 731
0Y 699 GRGQVYSSGRPLKCYVVGAGELPDNFVDSGTGLTFSIPSPSEMYRMSLE 747
Db 732 gsmryfaskkpractldegedvfykldqdg-muvvqvr-----wpld 771

```

	RESULT	11
AA17417	ID	AA17417 standard; Protein; 784 AA.
XX		
AC	AA17417;	
XX		
D7	29-JUL-1999	(first entry)
XX		
DE	Cucumber raffinose synthase.	
XX		
KW	Raffinose synthase; sucrose; galact	
XX		
OS	Cucumis sativus.	
XX		
PN	JPI123080-A.	
XX		
PD	11-MAY-1999.	
XX		
PF	24-OCT-1997;	97JP-0292969.
XX		

PD	11-MAY-1999.
XX	
PF	24-OCT-1997;
XX	97JP-0292969

XX Sequence 783 AA;

Query Match 34.5%; Score 1379; DB 20; Length 783;
Best Local Similarity: 39.1%; Pred. No. 1,4e-133;
Matches 295; Conservative 126; Mismatches 279; Indels 54; Gaps 16;

```
QY 5 PKISVNDGKLVHGGKTIITGVPDNVVLTGSGRGLVTGAEVGAATASHSKSLHVEPKVLE 64
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db pIIsEEnfmgvnhvlIsqVpsnItaSkmg---fdglfvgIdapepkarhvsyqglk 82

QY 65 CLRFECCEFFLMMWTQMGCGRDVPLETQFMLESKESETDGNSPIITYTVLLPLEG 124
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db gIpfmsIftrfwWtltwtsgnGrdeheqIlliIdksd---egIgrp--yIvIlplIeG 137

QY 125 QFRAVLQ-GNDKNEIEICLESNDNAVETDQGIHMVYMHAGTNPFEVINOAVKAEKHMQT 183
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db pfrasIqpgsvdydIcvesgstckvgsdfravlyIragpdpfkIkdtmkvqahlyt 197

QY 184 FLHREKKRLPSCLDWFGMCYDAETDYTAGVEBGLKSLSGGTPPRFLIIDGWOQIE 243
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db fKIldckIprgIvdKfgwtcdafyIkvepygvegvKglvengvppglvIlddgwqslc 257

QY 244 NKAQATKTCIVOE-----GAQFATRLTGIKENTKFKQKLONN-----EQMSGKHLV 290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db hdddpIdt---qegInrtsageqmpcrIkYeenktrdyksplmgheDpmmgrafv 314

QY 291 HGAKOH-HNVKNVYVMAHLAGVWGKVPATGMEHYDTALAYPVQSPGVLTGNQPDIVMS 349
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db rIkkeefktvehvywhafItygwsgvrpnvpglp--eaqvtvpkIsplIemtmEdIavdk 372

QY 350 LAVHGLVHPRKVFNFNEHAYLASGVGVKYDVONIIETLGAGHGRVSLRSTYHH 409
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db IvnngIglvppkagelIyegIhshlencgIdgvkvdvIhIlIemnaedygrvelaktyk 432

QY 410 ALEASIASNFTDNGCIACGCHNTDGLYSAKQT-AIVRASDDEYPRDPAS----- 457
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 433 altesvrkhfkngvIaamegndfmlIgtetIcIgyrgddftwptcdpsgdIngtywIggc 492

QY 458 HTHTISSVAYNSLFLGEFMQPDWDMFHSILHPADYHAARAIGCPIYVSDKPGNHFDL 517
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 493 hmvh---caynsImgnfIhpdmfsgtchpcaefhaasraIsqgplYvsdvvgkhnIrl 549

QY 518 LKKVLVPOGSVLRADLPGRPTRDSLFDVPARDRISLKIWNLNKCSGVGVFNCOGAGWC 577
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 550 lkrIvIadgsIlrceyhaIptcdclfvdpIhdgkcmIkImnInkYngvIyvfncqggsws 609

QY 578 KIEKTRRIDTSPGTLTASVCASDVDLIT-----OVAGAEMLGDTIVAYRSGEVIRLPK 632
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 610 resrkhIctseykskpsIctskpKdewenghkpfIkYvecfa---myftkexkIlIsqI 666

QY 633 GVSIPVTLKLEFELEFHCPIQELAP--SISPAIGLDMENTGGAVEQVEIHNRPAATKT 690
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 667 sdtIeIsIdpfdyellIvpspm-tIlpwesIafapIglvImInaagavksIdIsednekM 725

QY 691 IALSVGRGRFGVYSSQRPLKCVGGAETDENYD 724
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 726 vqvgIkagemmvyseKpkacrIvngedmeIeye 759
```

Search completed: June 4, 2002, 09:06:11
Job time: 599 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 4, 2002, 08:55:37 ; Search time 14.71 Seconds
(without alignments)
1245.357 Million cell updates/sec

Title: US-09-810-186-1

Perfect score: 3996
Sequence: 1 MVTLPKISVNDGKLVHGKT.....TFISPYSPREMYRMSIEIQV 750

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA.*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

. Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1394	34.9	784	US-08-846-234-5	Sequence 5, Appl1
2	119.5	3.0	1074	US-09-004-838-111	Sequence 111, App
3	117.5	2.9	4544	US-08-469-486-52	Sequence 52, Appl
4	117.5	2.9	4544	US-08-469-658-52	Sequence 52, Appl
5	115	2.9	478	US-09-004-838-51	Sequence 51, Appl
6	109.5	2.7	780	US-08-232-538-14	Sequence 14, Appl
7	109.5	2.7	780	US-08-786-164-14	Sequence 14, Appl
8	109.5	2.7	1338	US-08-750-141A-3	Sequence 3, Appl1
9	109	2.7	477	US-09-004-838-46	Sequence 46, Appl
10	109	2.7	477	US-09-004-838-99	Sequence 99, Appl
11	106	2.7	30	US-08-846-234-1	Sequence 1, Appl1
12	103.5	2.6	587	US-08-931-608A-4	Sequence 4, Appl1
13	102	2.6	1817	US-09-004-838-125	Sequence 125, App
14	102	2.6	2206	US-07-852-260-2	Sequence 2, Appl1
15	102	2.6	2206	US-08-461-503-2	Sequence 2, Appl1
16	102	2.6	2206	US-08-465-250-2	Sequence 2, Appl1
17	100	2.5	758	US-08-874-678-1	Sequence 1, Appl1
18	100	2.5	758	US-08-643-839-1	Sequence 1, Appl1
19	100	2.5	758	US-08-051-363-24	Sequence 24, Appl
20	99.5	2.5	566	US-08-484-993B-41	Sequence 41, Appl
21	99.5	2.5	566	US-08-484-158B-41	Sequence 41, Appl
22	99.5	2.5	566	US-08-484-596A-41	Sequence 41, Appl
23	99.5	2.5	566	US-08-480-150A-41	Sequence 41, Appl
24	99.5	2.5	566	US-08-458-731-41	Sequence 41, Appl
25	99.5	2.5	566	US-08-149-223A-41	Sequence 41, Appl
26	98.5	2.5	1311	US-08-340-011-5	Sequence 5, Appl1
27	98.5	2.5	1311	US-08-901-710-5	Sequence 5, Appl1

28	98	2.5	750	6	5457037-3	Patent No. 5457037
29	98	2.5	1477	1	US-08-038-682-4	Sequence 4, Appl1
30	98	2.5	1477	1	US-08-302-832-4	Sequence 4, Appl1
31	98	2.5	1477	2	US-08-530-198-4	Sequence 4, Appl1
32	98	2.5	1477	2	US-08-469-880-4	Sequence 4, Appl1
33	98	2.5	1477	2	US-08-728-470-4	Sequence 4, Appl1
34	98	2.5	1477	2	US-08-617-697-4	Sequence 4, Appl1
35	98	2.5	1477	2	US-08-719-641-4	Sequence 4, Appl1
36	96.5	2.4	4303	2	US-08-460-751-2	Sequence 2, Appl1
37	96	2.4	776	4	US-09-346-237-4	Sequence 4, Appl1
38	94.5	2.4	610	2	US-08-974-565C-9	Sequence 9, Appl1
39	94.5	2.4	610	3	US-09-255-748-9	Sequence 9, Appl1
40	94	2.4	751	6	5457037-5	Patent No. 5457037
41	94	2.4	776	4	US-09-346-237-7	Sequence 7, Appl1
42	94	2.4	1004	4	US-09-268-347-30	Sequence 30, Appl
43	94	2.4	4302	3	US-08-658-136-5	Sequence 5, Appl1
44	93.5	2.3	610	2	US-08-942-521B-7	Sequence 7, Appl1
45	93	2.3	1198	4	US-09-199-637A-405	Sequence 405, App

ALIGNMENTS

```
RESULT 1
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-846-234-5

Query Match 34.9%; Score 1394; DB 4; Length 784;
Best Local Similarity 37.6%; Pred. No. 5,9e-146;
Matches 289; Conservative 143; Mismatches 285; Indels 52; Gaps 15;

OY 8 SVNDGKLVHGKTLITGVDPNVVLTGSGRGL-----VTGAPVGAATASAKSLHVPKGV 62
DB 26 AIDGSDFTVGHGSLFDLDPNIVASPYTSIDKSPVSGVGFQFDPASDPDSRHVAVSICK 85
OY 63 LELGRFMCERFLMWTOMTGCTGRDVPLETFQFMLEKSESEFDGNSPIIYTVLPLL 122
```

[illegible]

```

      FILING DATE: 09-JAN-1998
      CLASSIFICATION: 800
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/781,734
      FILING DATE: 10-JAN-1997
      ATTORNEY/AGENT INFORMATION:
        NAME: Einhorn, Gregory P.
        REGISTRATION NUMBER: 38,440
        REFERENCE/DOCKET NUMBER: 023070-078B10US
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: (415) 576-0200
        TELEFAX: (415) 576-0300
      INFORMATION FOR SEQ ID NO: 111:
      SEQUENCE CHARACTERISTICS:
        LENGTH: 1074 amino acids
        TYPE: amino acid
        STRANDEDNESS:
        TOPOLOGY: linear
      MOLECULE TYPE: protein
      FEATURE:
      NAME/KEY: -
      LOCATION: 1..1074
      OTHER INFORMATION: /note= "R62K deduced sequence"
US-09-004-838-111

Query Match          3.0%; Score 119.5; DB 4; Length 1074;
Best Local Similarity 18.9%; Pred. No. 0.0023;
Matches 130; Conservative 94; Mismatches 261; Indels 203; Gaps 29;

QY   133 NDKNEIETICESGDNVAVETDGLIHVVYHNA-GTNPFEVIINQAVKVEKNMOTFLH----- 186
DB   150 NDFESRETFPKALTEALISNHTSHVVALMGMGVGKTTMKRLKNIIEKRFENHVLV 209
QY   187 -REKRILRSCIDPMFCMTWDAFYTYD-----TAEGVEGSLTSQGSGTPREFLLID 236
DB   210 IKENNDDLSIQDAV-----ADYLDMKLTESNESERAKLRGFPARKDGKRRRLIID 263
QY   237 DGMQOIE-----NKADATECLVOEGAQAFATRLTGIKENTKFQKKLONNEQMSGL 286
DB   264 DWMOGVNMEDIGLSPFGQGVDFKVLITSENKDVCAM-KGEANLIPVKELTEEBAOSL 322
QY   287 KHLVGAQHNNHVKNVYWHNALAGYWGGV-----KPAATGMEHD- 326
DB   323 FYQPVKSDTHLDK--IGKALVRNGGGLPIAIIKTIAATLKRNKRDVMKDALSRLEHHDI 379
QY   327 TALAAVY-----QS----PGVLGNQPDIVNDSLAVHGLGLVHPRKVEN--- 365
DB   380 ELIAIVVFQMSYDNLQNNEAQASIFLCGLPREDFDIPPEELVRYGMGL---RVENGYYT 435
QY   366 ---FNEIHAYLIASGVGVKVDVONIETLTGAGHGGRVSILFRSYHNHLEASIASNFT-- 420
DB   436 IGEARHRLNAVITYELLKDSNLLIESDV-----HCIRM-----HDLVRAVLDTFNRF 482
QY   421 -----DNGIACMCNHTDGLYSAKOTALVARASDEVPRPDPASHTHISSVAANSYSLF 472
DB   483 KHSLLVNVNGSGMLGWPRMDMSASCCKRISSLCKMSDFPRD-----VKRPNNLLI 532
QY   473 GEFMQPDMDMFHSLRPADYHAARAAGCGRIYVSDKRGNNHFOLKILVLPDGSVLAHQ 532
DB   533 LKIMHAD---KSLEFPDDFFYEMKKLOVI-----SYDHMKYPLPTS----- 571
QY   533 LPGRRTROSLFPPARDRTSILKILMNKKGSGVGVGFNQOGAGWKCIIEKTRPHDTSPT 592
DB   572 -PQGSTN---r-----LRYVJLHQC--LMFDCSSLG-----NL 598
QY   593 LTASYCASDVDLITOVAGAEMWLGDIT-----YYAYSSEGIYIRLPKGVSIPLVTLVLEF 645
DB   599 INLEV-----LSFANGSIGEMPLSTIGNLKELRYDLTLNCODGLRIDNG---VLKTKVL 648
QY   646 ELTFHCPIDQELAPSLISFAIIGLLDMFNFGAVEQYEIH--NRATKI-----IA 692
DB   649 EELVRYVGGRYOKALSTFDENCENCAERSKMLSALEEFEEFKNNAPRKMMSENEREKRIS 708

```

Query Match	2.9%;	Score 117.5;	DB 2;	Length 4544;
Best Local Similarity	20.8%;	Pred. No. 0.057;		
Matches	98;	Conservative 51;	Mismatches 178;	Indels 145;
				Gaps 25;

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,538
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Wallen, John W. III
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 188881A
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 780 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-232-538-14

Query Match 2.7%; Score 109.5; DB 1; Length 780;
Best Local Similarity 17.7%; Pred. No. 0.017;
Matches 111; Conservative 82; Mismatches 211; Indels 223; Gaps 25;

QY 86 CGRDVPLETQFMLESKESETDGENSPFIYTVLLPRLLEGOFRAYVAGNDKNEIICLESQ 145
DB 252 CTATPPLNTFRQMTWSYRDEKNKRAS-----VRRRIDQSNHANIIFYSVLTI 298
QY 146 DNAVETQOGLHMYMNGATNPFYINQAVKAEKHMOTFLHREKRLPSCIDMFGMCTMD 205
DB 299 DKMNKKDKGLTYCVRSGPS-FKSVNTSVHIYDKAFITYVHKKQOYLE----- 345
QY 206 AFYTDVTAEGVEEGLKSLSGGTP-PRFLIIDGWOQIENKA-----KATECL 253
DB 346 -----TVAGKRSYRLSMKYKAFPSPEVWMLKDGLPATEKSANYLRGYSLLIKDYTE-- 397
QY 254 VOEGAQFATRLTGKENTKQOKLQNNQSGKLHLYHGAKQHNKYNVYVHALLAGYNG 313
DB 398 --EDAGNTYLLISIKQSNVF-----KNLT--ATLIVNKPQIYEKAVSSF-- 438
QY 314 GVKPA--ATGMEHYDTALAYPVQSPGV-----LGNQPDIVMDSLAV 352
DB 439 -PDPAIPLIGSRQILCTAIGIPQRTIKMFWMHPCNNHSHARCDFCSNNDEESFLIDADSN 497
QY 353 HG-----LGLVHKKVYFNFNELHAYL--ASCGVDGVKVDY-QNIITFLGAG----- 396
DB 498 MGNRIESTITORMALIEGK-----NKMASPLVYADSRISGIVICIASNMKVGVGRNISFY 551
QY 397 ---HGGVSL-----TRSYHNL----- 411
DB 552 ITDVPNGFHVLEKMPTEGEDKLKSCVYKFLRDYTWILLRTVNNRTHMYSISKMAI 611
QY 412 --EASISNFT-----DNQCIACQCHNT--DGLYSAKQAIIVRASDEFYPRPASHT 459
DB 612 TKRESITLNTLIMVSLQDSGTACARANNYTGEBEILQKKEITIRQGEAPYLLRNLSDDT 671
QY 460 IHISVAANSLFLGEPQPDMDHSLHPADYHAARAIGCGPIYVSDPKGNHNDLLK 519
DB 672 VAISSTLTDCHANGVPEPQITFKNNH-----IQDEPG----- 706
QY 520 KLVLPDGSVLRALPGRPTRDSLEVDPAKDRSTSLKIIMLNKCSGVVGFNCGAGMCKI 579
DB 707 -IILPGS-----STLFIERYTEDE-----GVYHCKAT----- 734
QY 580 EKKTRIHDTSPGTLJASVCAVDVLT 606
DB 735 NOKGSVESSAYLVQGTSDKSNLELIT 761

RESULT 7
US-08-786-164-14

Sequence 14, Application US/08786164

Patent No. 5861484
GENERAL INFORMATION:
APPLICANT: THOMAS, KENNETH A.
APPLICANT: KENDALL, RICHARD L.
TITLE OF INVENTION: INHIBITOR OF VASCULAR ENDOTHELIAL
CELL GROWTH FACTOR
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000, 126 E. Lincoln Ave.
CITY: Rahway
STATE: NJ
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Microsoft Word 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/786,164
FILING DATE: 21-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mark Hand, J
REGISTRATION NUMBER: 36,545
REFERENCE/DOCKET NUMBER: 18888DA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-594-3905
TELEFAX: 908-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 780 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-786-164-14

Query Match 2.7%; Score 109.5; DB 2; Length 780;
Best Local Similarity 17.7%; Pred. No. 0.017;
Matches 111; Conservative 82; Mismatches 211; Indels 223; Gaps 25;

QY 86 CGRDVPLETQFMLESKESETDGENSPFIYTVLLPRLLEGOFRAYVAGNDKNEIICLESQ 145
DB 252 CTATPPLNTFRQMTWSYRDEKNKRAS-----VRRRIDQSNHANIIFYSVLTI 298
QY 146 DNAVETQOGLHMYMNGATNPFYINQAVKAEKHMOTFLHREKRLPSCIDMFGMCTMD 205
DB 299 DKMNKKDKGLTYCVRSGPS-FKSVNTSVHIYDKAFITYVHKKQOYLE----- 345
QY 206 AFYTDVTAEGVEEGLKSLSGGTP-PRFLIIDGWOQIENKA-----KATECL 253
DB 346 -----TVAGKRSYRLSMKYKAFPSPEVWMLKDGLPATEKSANYLRGYSLLIKDYTE-- 397
QY 254 VOEGAQFATRLTGKENTKQOKLQNNQSGKLHLYHGAKQHNKYNVYVHALLAGYNG 313
DB 398 --EDAGNTYLLISIKQSNVF-----KNLT--ATLIVNKPQIYEKAVSSF-- 438
QY 314 GVKPA--ATGMEHYDTALAYPVQSPGV-----LGNQPDIVMDSLAV 352
DB 439 -PDPAIPLIGSRQILCTAIGIPQRTIKMFWMHPCNNHSHARCDFCSNNDEESFLIDADSN 497
QY 353 HG-----LGLVHKKVYFNFNELHAYL--ASCGVDGVKVDY-QNIITFLGAG----- 396
DB 498 MGNRIESTITORMALIEGK-----NKMASPLVYADSRISGIVICIASNMKVGVGRNISFY 551

Qy	397	----	HGRVSL-----	----	TSYHNL-----	411
		:	1:1	:	1:1:1	
Db	552	ITDVPNGHVALEKMPTEGEDKLKSCYVKNFLYDVTWILLRTYNNNTMTYSISKOMAI	61111			
Qy	412	--EASIASNFT-----	DNMGCIACMCHNT--	DGLYSANOTAIVRASDDPYPRDPA	SHT	459
		:	1:1:1	:	1:1:1	
Db	612	TKHSITFLNTLTMVNSLODSSGTCYCARNAVYTGEEIIÖKKEITITRDEAPFLTNLSDHT	67171			
Qy	460	THHSVAANSILFLEFMOPMDMDSHSLHPADYHAAARAIGCGPIYVSDRKGNNFDLLK	519			
		:	1:1:1	:	1:1	
Db	672	VAISSSTLDDCHANGVPEPÖITWRKNHKK-----	IQÖEFG-----	706		
Qy	520	KLVLPDGSVLRAOLPGRFTRDLSLFVDPARDFTSLIKIWNLNKSGGVYVENCÖGACGMCKI	579			
		:	1:1	:	1:1:1	
Db	707	TLILPGS-----	STLFIERYTEDE-----	GVYHCKAT-----	734	
Qy	580	EKKTRIHDTSPOGTLASVACASDVLLIT	606			
		:	1:1	:	1:1:1:1	
Db	735	NÖKGVSESSAYILVÖGÖSTDKSNLELIT	761			

RESULT 8
 US-08-750-141A-3
 Sequence 3, Application US/08750141A
 Patent No. 6011003
 GENERAL INFORMATION:
 APPLICANT: Charnock-Jones, David S.
 APPLICANT: Boccock, Christine A.
 APPLICANT: Sharkey, Andrew M.
 APPLICANT: Smith, Stephen K.
 TITLE OF INVENTION: FLT-4-(FMS-LIKE TYROSINE-KINASE),
 TITLE OF INVENTION: FLT-15, VARIANTS THEREOF USED AS GROWTH FACTOR INHIBITORS
 NUMBER OF SEQUENCES: 17
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/750,141A
 FILING DATE: 12-MAY-1997
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 1396-1-001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 138 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 US-08-750-141A-3

Query Match	2.7%;	Score 109.5;	DB 3;	Length 1338;
Best Local Similarity	17.7%;	Pred. No. 0.046;		
Matches 111;	Conservative 82;	Mismatches 211;	Indels 223;	Gaps 25;
QY	86	CGRDVPLETQFPLETSKESSETDGENSPIITVYLLPLLESGQFRAVYQGDNDKNFEICLES	145	

Db	252	СТАПРЛНТНРОЖМВSRPDEKNNKRAS-----VRRIDDSNSHANIFYSLTI	298
Qy	146	DNAVEITDOGLHMYMAGTNPREFEYINQAVAKAVENKHOITFLNHEKKPLRPSLDMFGSTMD	205
Db	299	DKMONKCKRGITLCTVRBGRPS-FKSVNTRVHIYKAFITVHNRKOQVLE-----	345
Qy	206	AFYIDVTAECEVEEJLKSLSQGTPE-PRFLIJDGWOQIENKA-----KQATECL	253
Db	346	-----IVAGKRSRRLSMKKVKAAPSPREVLWYMKOLRPAETKSAHYLRGSLIKEVDYE--	397
Qy	254	VOEGAOEATRLITGKENTKROKJLQNNQOMSGKLHLVHAKQONHNKYNVYMHALLGYMG	313
Db	398	-----EAGANTYITLLSIKOSVNF-----KNLT--ATLIYVAKPQIYEKAVSSF--	438
Qy	314	GVKPA--ATGMEHEDFLALAYVPOSPGV-----LGNOPDIYMDSLAV	352
Db	439	-----PDRLALYLGSRQIITLCATVIGPRLIKWHPNCSNNHNSHSEARDCFSNNEESFILLADSN	497
Qy	353	HG-----LGLVHKKYVNFNENLHAYL--ASCQVDQKVDY-QNILETIGAG-----	396
Db	498	MGNRIESTITORMALIEEG-----NKMASITLVADSRISGIYICIASNKKVGTGRNISFY	551
Qy	397	-----HGGRVSL-----TRSYNHAL-----	411
Db	552	ITDVGNGHVHNEKMPTEGDEJLKSCTYNNKFLYRDVTWILLRVNNTNMYHSISKOMAI	611
Qy	412	-----EASIASNFT-----DNCSIAQMCCHNT--DGLYSAKOTALYRASDDFYPRDPASHT	459
Db	612	TKHEHITINTLITMNVSLDQSGTVCARANNVYTGEELLQKEITRLRDEAPRLILRNSDHT	671
Qy	460	IIHSSVAVANSLEJBEFMOQMDMFHSLRPADYHAARAALGGCPIYVSDRPGNHPEJLJK	519
Db	672	VAISSSTLLDCHANGVPERQITWKRNNHK-----IQDEBG-----	706
Qy	520	KLVLPDGSYLRAQILGRPTRDSLFPVDPARDTSILKILWNLKSGCVYGVYGRQAGAGCKI	579
Db	707	-----IILPGS-----STLFIERTEDE-----GVYHCKAT-----	734
Qy	580	EKKTIHDSPTGLTASVACASVDVIT	606
Db	735	NQKGSVESSAVITLVQIGDSKSNLEIIT	761

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Query Match	2.7%;	Score 109;	DB 4;	Length 477;
Best Local Similarity	18.4%;	Pred. No. 0.0076;		
Matches 79;	Conservative 62;	Mismatches 147;	Indels 142;	Gaps 17

RESULT 10
US-09-004-838-99
Sequence 99, Application US/09004838
Patent No. 6350533
GENERAL INFORMATION:
APPLICANT: Michelmoe, Richard W.
APPLICANT: Shen, Kathy
APPLICANT: Meyers, Blake
TITLE OF INVENTION: Procedures and Materials for
TITLE OF INVENTION: Confering Pest Resistance in Plants
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor

Query Match	2.7%;	Score 109;	DB 4;	Length 477;
Best Local Similarity	18.4%;	Pred. No. 0.0076;		
Matches 79;	Conservative 62;	Mismatches 147;	Indels 142;	Gaps 17;

```

QY 213 AEEVEGLSTLSOGTTPRELLIDDMGMOLE-----KKAQATCCLVQEAQAT 26
Db 57 AADLRQGFQKDSKDGKKNKFVYLDDVWQSVLEDIGLSPPNQGVDFKYLTSRDRHYCT 116
QY 263 RLLGKENTKRFQKQLONNEQMSGLKH--LYHGAKOHNN-----YKVVY----- 303
Db 117 -YGVGEAKLILWGLLLEAEQSLFHFQFVVTSEPELHKIGEDLYVKKCFELPILAIKTACT 177
QY 304 VMIALGYGCGYKPRATGMENTALARP-----VOSPEVLGNOPDI 345
Db 176 LRKRRKDAW---KDLASLELHHDIGQSVVPKVFETSYNNLKDEKTSVFLMGCLFPEEDI 232
QY 346 VMDSLAVHGLGVHKKYKFNFYNEL-----HAYLASCQDGVKQVONII 396
Db 233 PIELMRKYOMGL---RLFDKVNITITQARRKRLTCIELRYVHNLLESYDGVHYKMHDLV 288
QY 391 ETLGAGHGGRSVLTSRHYHNALEASIASN---FTDNGCIACMCHNTDGLYSANQOTIVR 445
Db 289 RAVFLVMEFESEV-----HASIVHGNNPMPTEN-----DMTDSCKQISLTC 322
QY 446 ASDDFEPRPASPHTIHLSVAYNSLFLGEMQODMWFHSLHPADYHAARAALGGCPY 505
Db 330 KSMLEFPDGLKRPNIKLITKLHNG---GKSLRYPQFVQGM----- 366
QY 506 VSDKPRENNHFDLLKLYLPDGSVYLRAOLFPRPRDSLFLVPDARDTSLKLTWNLNKCSGV 565
Db 367 --EKLEVISYDEMKYPLRP-----SLP-----QGSTILRYLVHLHRECS-- 401
QY 566 VGVFNCQCGAG 575
Db 402 LRFDFCDSSTIG 411

```



```

NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-078810US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 125:
SEQUENCE CHARACTERISTICS:
LENGTH: 1817 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: 1.1817
LOCATION: 1.1817
OTHER INFORMATION: /note= "RG2S deduced sequence"
US-09-004-838-125

```

```

Query Match      2.6%; Score 102; DB 4; Length 1817;
Best Local Similarity 18.1%; Pred. No. 0.56;
Matches 141; Conservative 105; Mismatches 266; Indels 268; Gaps 37;

OY 58 FPMGLLEGIRFMC--FRFL-----MMTQRMGTCGRD-----VPLETQFMLE 100
DB 89 FPIDVI-----TCSLRIRHKLQKRAFKITQIESLTRQLSLISWTDPPVBLGR---VG 139
OY 101 SKESETDENSEPIIYTVLLPLLEGFRFVAVLOGNDKNEIEICLESQDNVAVETDQGLHWYV 160
DB 140 SMNASTSASSSD-----FPSREKFTQALK-----ALENNQGFHWAL 178
OY 161 --HAGTNEFEVYNQAVKV--EKHMOTFLHRE---KKRLPSCLE---DMFGMCTWDAFYT 209
DB 179 CGMGVGVGTRKRMQRKKAERKLFNFYIRAVIGEKTDPAIOBAIADYLS-IQINERTK 237
OY 210 DVTAGVEEGLKLSQGGTPPRFLIIDCGMOIE-----NKAKDAIECLVQBSAQ 259
DB 238 PARADKLEMEFKKNSDGKTFLLYLDVWQVLDIEDIGLSPFNQGVDFVLLTSRSQ 297
OY 260 FATR-----LTGIKENTKFOKKLQNN-----QMSGKLHLVGA 294
DB 298 VCTMVGVEANSIIINGGLLFEAQSLEFQGFVETSEPELOKIGEDIVKCCCLPAITMA 357
OY 295 QHNHNKYNVYVHALAGYGVKPRATGMEHTALAP-----VQSP----- 336
DB 358 CTLNKKRKDAW-----KDALSRIEHYDINHVA PKVETSYHNLQEEETKSTFLMC 407
OY 337 GVLGNQPIVMDSLAVHGLGLVHPKVFENFYNELHAVLASGVGVAVDVONIIETLGAG 396
DB 408 GLFPDPDPIPEELMKRYOMGLKIDRYVTT-REARTRLNTC-----IERLVOT----- 454

```

```

OY 397 HGRVSLTRSYHHALEASIASNFITDNGCIACMCHNTDGLYSAKOTAVRASDDFYPRDPA 456
DB 455 -----NLLIESDDVGCY-----KMHDLVRA-----FVLGMF 480
OY 457 SHTIISSAVNSLFLGFMQPMDMFSLRPADYHRAAARIGCPIYYSKDKGNINFD 516
DB 481 SEVENASTVNHGNM-----PEW-----TENDITDSCRISLTCKSMKFPDGRFP 526
OY 517 -LLKRLVLPDGSVLRAD--LPGRPTRDSLFD-----PARDTSL-LKIMNLKNC 564
DB 527 NLMILKMHGKSLRFPDGFEGMEKLVISYDKMKYPLPLARCSNTINVLHLTCS- 585
OY 565 VGVFNCQAGWCCKIEKKTRIHTSPGTLTASVCAVDLITQYAGA-----EWLGDTI-- 618
DB 586 -LKMFDGS-----CIGNLSNLEVLSFANSRIEMLPSTVRN 619
OY 619 -----VAVRSGEYIRLPKGSIPVTLKYLEFELFHCPCPIDEIAPISFALGLD----- 669
DB 620 LKRLRLDLPFCDGLRLEQGV-LKSLVLEEFYIGN-----ASGFIDNCN 664
OY 670 -----MFTGGAVEQVEIHNRAATR-TTALSVRGRGRFGVYSSORPLKCV 713
DB 665 EMARSDNLSALEFAFPNNKALEVKNMSEFNLERKISVGRSFDQNTIMSSHVENMLQLY 724

```

```

RESULT 14
US-07-852-260-2
Sequence 2, Application US/07852260
Patent No. 5525715
GENERAL INFORMATION:
APPLICANT: Racanietello, Vincent
APPLICANT: Tatem, Joanne M.
APPLICANT: Weeks-Levy, Carolyn L.
TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES FROM
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/852,260
FILING DATE: 19920619
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 36607-B-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2206 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-852-260-2

```

```

Query Match      2.6%; Score 102; DB 1; Length 2206;
Best Local Similarity 21.0%; Pred. No. 0.8;
Matches 122; Conservative 67; Mismatches 207; Indels 184; Gaps 33;

```

```

QY 10 NDGLVYH-GKTLIGVDPNVLTFRG-----SGR-----GLV 40
DB 1660 NDGLVLTSTSKYPMNMYVPGAVTEOGYLNLDGKOTARILMNEPTIRAGCGGVTTCGKY 1719
QY 41 TGAFTGATASHSKSLHVFPMGVLEGLRFMCCEFRKLMMTORMTGCGRDVPLETFMLIE 100
DB 1720 IGMHVGNGSH-----GFAAL-----KRSYFTQSOG-----EIOMNR-P 1753
QY 101 SKESETDGENSPITTYVLLPLE-GOPRAVLQGNCKNEIEICLESQDNAVETDGLHMY 159
DB 1754 SKFA-----GYPLINAPTKTLEPSAFHYFEGVKEPAV---LTKNDPRLKTFDEEALFS 1805
QY 160 MHAGNPFVYINQAVKAVEKHMOTFLHREKRLPSCLDMFGMCTWDAFY--TDVTAEGVEE 218
DB 1806 KYVGKITEVDEYMEKAVDHYAGOLMSLIDISTEQMCLF-----DAMGTID-----1850
QY 219 GLKSLF--SOGTTPRFLIIDGMOQIENK---AKDATECLVQEGAOFATRLTGIKENT 271
DB 1851 GLEALDLSSTAGYP--YVAMGKKKKRDLINKOTRDTKEMORLIDAYGINPL-VTVYKDEL 1907
QY 272 KFOKTLQNNQMSGKHLVHGAKQHNKANYVMHALLAGVGVKRAATGMEHYDTALAY 331
DB 1908 RSKTYE-----QKSRLEIASSLINDSV-----AMRAFGNLYAAF---H-----1944
QY 332 PYOSPGV-----IGNODIYMDSLAVHGLGVHPKVFNF-YNELHAYLASCGVDGKYD 385
DB 1945 --RNGVYVTSAGVCDPDLFWSKIPY-----LMEKLFADYTGVDASLSPAMFEALKMY 1997
QY 386 VONITETLAGHGRVY---LTRSYHHALEASIASNFTDNGCIACMCHNTDGLYSAKOT 441
DB 1998 LEKI-----GFGDRVDYIDYLNHS--HHLYKNKI---YCVKGMPSGSGSTIFNSMINN 2047
QY 442 AYVRA-----SDDF---YPRD-----PASHTIH 461
DB 2048 LIITLLKTYKIGIDLHKMIAYGDDVYASTPHEVDASLSLAOSKDYGLTMTPADKSAT 2107
QY 462 ISSVAY-NSLFLGEFMOPMD---MFHSLHPADYHAAR 497
DB 2108 FEYTWENTVFLKRFRADEKYPFLIHPVMPKMEIHESIR 2147

RESULT 15
: Sequence 2, Application US/08461503
: Patent No. 5834302
: GENERAL INFORMATION:
: APPLICANT: Racanietello, Vincent
: APPLICANT: Tatem, Joanne M.
: TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES
: TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/461,503
: FILING DATE: 5-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 36607-D-PCT-US
: TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: TELEFAX: 422523 COOP UT
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2206 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-461-503-2

Query Match 2.6%; Score 102; DB 2; Length 2206;
Best local Similarity 21.0%; Pred. No. 0.8;
Matches 122; Conservative 67; Mismatches 207; Indels 184; Gaps 33;

QY 10 NDGLVYH-GKTLIGVDPNVLTFRG-----SGR-----GLV 40
DB 1660 NDGLVLTSTSKYPMNMYVPGAVTEOGYLNLDGKOTARILMNEPTIRAGCGGVTTCGKY 1719
QY 41 TGAFTGATASHSKSLHVFPMGVLEGLRFMCCEFRKLMMTORMTGCGRDVPLETFMLIE 100
DB 1720 IGMHVGNGSH-----GFAAL-----KRSYFTQSOG-----EIOMNR-P 1753
QY 101 SKESETDGENSPITTYVLLPLE-GOPRAVLQGNCKNEIEICLESQDNAVETDGLHMY 159
DB 1754 SKFA-----GYPLINAPTKTLEPSAFHYFEGVKEPAV---LTKNDPRLKTFDEEALFS 1805
QY 160 MHAGNPFVYINQAVKAVEKHMOTFLHREKRLPSCLDMFGMCTWDAFY--TDVTAEGVEE 218
DB 1806 KYVGKITEVDEYMEKAVDHYAGOLMSLIDISTEQMCLF-----DAMGTID-----1850
QY 219 GLKSLF--SOGTTPRFLIIDGMOQIENK---AKDATECLVQEGAOFATRLTGIKENT 271
DB 1851 GLEALDLSSTAGYP--YVAMGKKKKRDLINKOTRDTKEMORLIDAYGINPL-VTVYKDEL 1907
QY 272 KFOKTLQNNQMSGKHLVHGAKQHNKANYVMHALLAGVGVKRAATGMEHYDTALAY 331
DB 1908 RSKTYE-----QKSRLEIASSLINDSV-----AMRAFGNLYAAF---H-----1944
QY 332 PYOSPGV-----IGNODIYMDSLAVHGLGVHPKVFNF-YNELHAYLASCGVDGKYD 385
DB 1945 --RNGVYVTSAGVCDPDLFWSKIPY-----LMEKLFADYTGVDASLSPAMFEALKMY 1997
QY 386 VONITETLAGHGRVY---LTRSYHHALEASIASNFTDNGCIACMCHNTDGLYSAKOT 441
DB 1998 LEKI-----GFGDRVDYIDYLNHS--HHLYKNKI---YCVKGMPSGSGSTIFNSMINN 2047
QY 442 AYVRA-----SDDF---YPRD-----PASHTIH 461
DB 2048 LIITLLKTYKIGIDLHKMIAYGDDVYASTPHEVDASLSLAOSKDYGLTMTPADKSAT 2107
QY 462 ISSVAY-NSLFLGEFMOPMD---MFHSLHPADYHAAR 497
DB 2108 FEYTWENTVFLKRFRADEKYPFLIHPVMPKMEIHESIR 2147

```

Search completed: June 4, 2002, 09:05:32
 Job time: 595 sec

Tue Jun 4 16:39:14 2002

us-09-810-186-1.rai

Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 4, 2002, 09:03:02 ; Search time 24.15 Seconds

(without alignments)
2984.141 Million cell updates/sec

Title: US-09-810-186-1

Sequence: 1

Sequence: 1 MTVPKISVNDGKLVHNGKT.....TFSIPSPDEMKRSIEIQV 750

Scoring table:

BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3219.5	80.6	773	2	T46188
2	3181.5	79.6	765	2	S45033
3	2951.5	73.9	757	2	S27762
4	1915.5	47.9	1170	2	C96599
5	1307	32.7	807	2	C85025
6	1228.5	30.7	357	2	T09530
7	753.5	18.9	371	2	T01717
8	442.5	11.1	649	2	D90496
9	126	3.2	657	2	B71367
10	125	3.0	982	2	S16292
11	119.5	3.0	1715	2	T30559
12	117.5	2.9	4544	1	S02392
13	115	2.9	928	2	C86546
14	115	2.9	928	2	C81591
15	114.5	2.9	1447	2	C86474
16	114	2.9	4545	1	S25111
17	111	2.8	953	2	F87466
18	110	2.8	905	2	D86380
19	109.5	2.7	1338	2	S09982
20	108	2.7	714	2	A40614
21	106	2.7	976	2	F81722
22	105	2.6	1601	2	AB1730
23	104.5	2.6	1127	2	T21635
24	104	2.6	754	2	T45910
25	104	2.6	1669	2	T49244
26	103.5	2.6	587	2	A55368
27	103	2.6	1408	2	H69068
28	102.5	2.6	386	2	D84787
29	102	2.6	452	2	T35762

ALIGNMENTS

30	102	2.6	1503	2	T43166	alpha-2-macroglobu
31	102	2.6	1813	2	T30564	resistance protein
32	102	2.6	2206	1	GNNY4P	genome polyprotein
33	102	2.6	2206	2	S03822	genome polyprotein
34	102	2.6	2207	1	GNNY5P	genome polyprotein
35	101.5	2.5	374	2	B82168	probable alpha-1,6
36	101.5	2.5	675	2	B96542	unknown protein [f
37	101.5	2.5	1158	2	AF1852	hypothetical prote
38	101.5	2.5	1430	2	D82533	RNA polymerase bet
39	101.5	2.5	1477	2	B43855	high-molecular-wet
40	101	2.5	1820	2	S71853	genome polyprotein
41	101	2.5	2209	1	GNNY2P	genome polyprotein
42	101	2.5	4466	1	S17231	dynein beta heavy
43	101	2.5	4466	1	S17653	dynein beta heavy
44	100.5	2.5	1171	2	T12956	hypothetical prote
45	100	2.5	949	2	E75352	glycine cleavage s

RESULT 1

Imbibition protein homolog - Arabidopsis thaliana
N:Alternate names: protein T8H10.120
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46188
R:Benes, V.; Rechmann, S.; Borckova, D.; Ansoorge, W.; Mewes, H.W.; Lemcke, K.; Mayer, submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23014
A:Accession: T46188
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-773 <BEN>
A:Cross-references: EMBL:AL133248
A:Experimental source: cultivar Columbia; BAC clone T8H10
C:Genetics:
A:Map position: 3
A:Insertions: 64/2; 146/2; 176/2; 192/3; 223/2; 259/2; 300/2; 484/3; 507/2; 552/2; 625/1
A:Note: T8H10.120

Query Match 80.6%; Score 3219.5; DB 2; Length 773;

Best Local Similarity 76.6%; Pred. No. 1,3e-240;

Matches 593; Conservative 73; Mismatches 83; Indels 25; Gaps 3;

QY	1	MTVPKISVNDGKLVHNGKTILGVPDNNVLTPOSGRGVTAFAVGATASHSKSLHVPFM	60
DB	1	MTIINSISVQNDNLYVGGKTIILKIPNNILLITVYTGNGFVSGSFGITGEQSKSLHVPF	60
QY	61	GVLEGLFPMCCFRKKMMWMTQRMGTGGRVPLETQFWLIESK-ESFDGENSPIIYTVLL	119
DB	61	GVLEGLFPMCCFRKKMMWMTQRMGTGGRVPLETQFWLIESK-ESFDGENSPIIYTVLL	119
QY	120	PLEGGQRAVLQGDNDKEITICLESQNAEFTDQGLHMYMAGCTNPEFYINQAVKVER	179
DB	120	PLEGGQRAVLQGDNDKEITICLESQNAEFTDQGLHMYMAGCTNPEFYINQAVKVER	179
QY	121	PLEGGQRAVLQGDNDKEITICLESQNAEFTDQGLHMYMAGCTNPEFYINQAVKVER	180
DB	121	PLEGGQRAVLQGDNDKEITICLESQNAEFTDQGLHMYMAGCTNPEFYINQAVKVER	180
QY	180	HMOTFLREKKRLLPSCIDMFGMCTWDAFYDYTAEGVEEGIKSLISOGGTPPRLLIIDGM	239
DB	180	HMOTFLREKKRLLPSCIDMFGMCTWDAFYDYTAEGVEEGIKSLISOGGTPPRLLIIDGM	239
QY	240	QOIEENKAKDTECLVQGAQFATRLTGIEKENTKFOKKLQNNQMSGLKHLVHGAKOHNV	299
DB	240	QOIEENKAKDTECLVQGAQFATRLTGIEKENTKFOKKLQNNQMSGLKHLVHGAKOHNV	299
QY	241	QOIEENKAKDTECLVQGAQFATRLTGIEKENTKFOKKLQNNQMSGLKHLVHGAKOHNV	299
DB	241	QOIEENKAKDTECLVQGAQFATRLTGIEKENTKFOKKLQNNQMSGLKHLVHGAKOHNV	299
QY	300	KNYVYVHALAGYGVGKPAATGMEHYDTALAYPVQSGVIGNPDIVMNSLVHGLGVN	359
DB	300	KNYVYVHALAGYGVGKPAATGMEHYDTALAYPVQSGVIGNPDIVMNSLVHGLGVN	359
QY	300	KNYVYVHALAGYGVGKPAATGMEHYDTALAYPVQSGVIGNPDIVMNSLVHGLGVN	359
DB	300	KNYVYVHALAGYGVGKPAATGMEHYDTALAYPVQSGVIGNPDIVMNSLVHGLGVN	359
QY	360	PKVVFNFYNELHAYLASCGVDGVKVDVQNIITLGAHGGRVSLTFRSYHHALEASIASNF	419
DB	360	PKVVFNFYNELHAYLASCGVDGVKVDVQNIITLGAHGGRVSLTFRSYHHALEASIASNF	419

Db 360 PKRVFNENLHSLYASCGIDGVKDVONIIETLGAGGGRVSLTNSYOOALFASIAKRF 419
QY 420 TDNGCIACMCHNTDGLYSAKOTAIVRASDDFYPRDPASHTIHISYAANSFLGFEKMPDW 479
Db 420 TDNGCIACMCHNTDGLYSAKOTAIVRASDDFYPRDPASHTIHISYAANSFLGFEKMPDW 479
QY 480 WDMFSLHPADYHAARAIGGCPITYSDKPGNHNDLKLKLVLPDGSVYLRQAOLGPRPTD 539
Db 480 WDMFSLHPAEYHAARAAGCAIYVSDKPGNHNDLKLKLVLPDGSVYLRQAOLGPRPTD 539
QY 540 DLFVDPARDRTSLKIMNLCSCGVYGVNCOGAGMCKIEKTRIHDTSPGTLASVCA 599
Db 540 DLFVDPARDRTSLKIMNLCSCGVYGVNCOGAGMCKIEKTRIHDTSPGTLASVCA 599
QY 540 DLFVDPARDRTSLKIMNLCSCGVYGVNCOGAGMCKIEKTRIHDTSPGTLASVCA 599
Db 540 DLFVDPARDRTSLKIMNLCSCGVYGVNCOGAGMCKIEKTRIHDTSPGTLASVCA 599
QY 600 SDVDLITOVAGAEMWLDITIVYARSGEVIRLPKGVSIPTVLKYLEELFHCPIQETIAPS 659
Db 600 SDVDLITOVAGAEMWLDITIVYARSGEVIRLPKGVSIPTVLKYLEELFHCPIQETIAPS 659
QY 600 SDVDLITOVAGAEMWLDITIVYARSGEVIRLPKGVSIPTVLKYLEELFHCPIQETIAPS 659
Db 600 SDVDLITOVAGAEMWLDITIVYARSGEVIRLPKGVSIPTVLKYLEELFHCPIQETIAPS 659
QY 660 ISPAIGLDMFNTGAVOEVEH-----NRATKTIATLSVR 696
Db 660 ISPAIGLDMFNTGAVOEVEH-----NRATKTIATLSVR 696
QY 697 GRGRFEGVYSSORPLKCVYGAETDFNYDSEGLTTFESIIPVSPBEMIRMSIEIQV 750
Db 697 GRGRFEGVYSSORPLKCVYGAETDFNYDSEGLTTFESIIPVSPBEMIRMSIEIQV 750
QY 720 GGRFEGVYSSORPLKCVYGAETDFNYDSEGLTTFESIIPVSPBEMIRMSIEIQV 750
Db 720 GGRFEGVYSSORPLKCVYGAETDFNYDSEGLTTFESIIPVSPBEMIRMSIEIQV 750

RESULT 2
S45033
probable imbibition protein - wild cabbage
C:Species: Brassica oleracea (wild cabbage)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C:Accession: S45033
R:Fujikura, Y.; Karszen, C. K.
submitted to the EMBL Data Library, May 1994
A:Description: Cauliflower cDNA encoding a putative imbibition protein.
A:Reference number: S45033
A:Accession: S45033
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-765 <FUU>
A:Cross-references: EMBL:X79330; NID:9488786; PID:9488787

Query Match 79.6%; Score 3181.5; DB 2; Length 765;
Best Local Similarity 76.5%; Pred. No. 1,1e-237;
Matches 587; Conservative 74; Mismatches 87; Indels 19; Gaps 3;
QY 1 MVTPEKISVNDGKLVVHGKTIITGVDPNVVLTGSGRGVLTGAVGATASHSKSLHVFPM 60
Db 1 MVTPEKISVNDGKLVVHGKTIITGVDPNVVLTGSGRGVLTGAVGATASHSKSLHVFPM 60
QY 61 GVLGELRFMCCEFRKLMWTOQRMGTGGRDVPLETOFMLEIESKESFTDGENSPIIYTVLP 120
Db 61 GVLGELRFMCCEFRKLMWTOQRMGTGGRDVPLETOFMLEIESKESFTDGENSPIIYTVLP 120
QY 121 LLEGFRVAVLQNDKNEIEICLSEGDNAVEEDOGILHMYMAGTNPFEVINOAVKAKEXH 180
Db 121 LLEGFRVAVLQNDKNEIEICLSEGDNAVEEDOGILHMYMAGTNPFEVINOAVKAKEXH 180
QY 180 MGTFFHREKKRLPSFVDMGCTWDAFYTDVTAEGVDGLSLSGGPRPRLIIDGMO 239
Db 180 MGTFFHREKKRLPSFVDMGCTWDAFYTDVTAEGVDGLSLSGGPRPRLIIDGMO 239
QY 241 QIENKAKDTECLVDEGAFATRLTGIEKENTFKOKLONNEDMSGLKHLVHGAQHNVK 300
Db 241 QIENKAKDTECLVDEGAFATRLTGIEKENTFKOKLONNEDMSGLKHLVHGAQHNVK 300
QY 301 NYVYHAIAGVGVKPRATGMEHDTALAYVOSPGVLTGNOPODIYVMSLAVHGLVHP 360
Db 301 NYVYHAIAGVGVKPRATGMEHDTALAYVOSPGVLTGNOPODIYVMSLAVHGLVHP 360
QY 361 KKVFNENLHSLYASCGIDGVKDVONIIETLGAGGGRVSLTNSYOOALFASIAKRF 420
Db 361 KKVFNENLHSLYASCGIDGVKDVONIIETLGAGGGRVSLTNSYOOALFASIAKRF 420

Db 359 KKVFNENLHSLYASCGIDGVKDVONIIETLGAGGGRVSLTNSYOOALFASIAKRF 418
QY 421 DNGCIACMCHNTDGLYSAKOTAIVRASDDFYPRDPASHTIHISYAANSFLGFEKMPDW 480
Db 421 DNGCIACMCHNTDGLYSAKOTAIVRASDDFYPRDPASHTIHISYAANSFLGFEKMPDW 480
QY 481 DMEFSLHPADYHAARAIGGCPITYSDKPGNHNDLKLKLVLPDGSVYLRQAOLGPRPTD 540
Db 481 DMEFSLHPAEYHAARAAGCAIYVSDKPGNHNDLKLKLVLPDGSVYLRQAOLGPRPTD 540
QY 541 DLFVDPARDRTSLKIMNLCSCGVYGVNCOGAGMCKIEKTRIHDTSPGTLASVCA 600
Db 541 DLFVDPARDRTSLKIMNLCSCGVYGVNCOGAGMCKIEKTRIHDTSPGTLASVCA 600
QY 599 DADLISEVAGQDMGDSIYAYARSGEVIRLPKGVSIPTVLKYLEELFHCPIQETIAPS 659
Db 599 DADLISEVAGQDMGDSIYAYARSGEVIRLPKGVSIPTVLKYLEELFHCPIQETIAPS 659
QY 601 SDVDLITOVAGAEMWLDITIVYARSGEVIRLPKGVSIPTVLKYLEELFHCPIQETIAPS 659
Db 601 SDVDLITOVAGAEMWLDITIVYARSGEVIRLPKGVSIPTVLKYLEELFHCPIQETIAPS 659
QY 661 ISPAIGLDMFNTGAVOEVEH-----NRATKTIATLSVR 703
Db 661 ISPAIGLDMFNTGAVOEVEH-----NRATKTIATLSVR 703
QY 704 YSSORPLKCVYGAETDFNYDSEGLTTFESIIPVSPBEMIRMSIEIQV 750
Db 704 YSSORPLKCVYGAETDFNYDSEGLTTFESIIPVSPBEMIRMSIEIQV 750
QY 719 YSSORPLKCVYGAETDFNYDSEGLTTFESIIPVSPBEMIRMSIEIQV 750
Db 719 YSSORPLKCVYGAETDFNYDSEGLTTFESIIPVSPBEMIRMSIEIQV 750

RESULT 3
S27762
Sip1 protein - barley
C:Species: Hordeum vulgare (barley)
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S27762
R:Heck, G.R.; Dorsett, C.; Ho, T.H.
submitted to the EMBL Data Library, February 1992
A:Description: Cloning and characterization of a gene, Sip1, associated with seed imb
A:Reference number: S27762
A:Accession: S27762
A:Molecule type: DNA
A:Residues: 1-757 <HEC>
A:Cross-references: EMBL:M77475; NID:91677099; PID:91677100
A:Genetics:
A:introns: 64/2; 151/1; 181/2; 197/3; 489/3; 557/2

Query Match 73.9%; Score 2951.5; DB 2; Length 757;
Best Local Similarity 71.1%; Pred. No. 6.6e-220;
Matches 540; Conservative 92; Mismatches 116; Indels 11; Gaps 5;
QY 1 MVTPEKISVNDGKLVVHGKTIITGVDPNVVLTGSGRGVLTGAVGATASHSKSLHVFPM 60
Db 1 MVTPEKISVNDGKLVVHGKTIITGVDPNVVLTGSGRGVLTGAVGATASHSKSLHVFPM 60
QY 61 GVLGELRFMCCEFRKLMWTOQRMGTGGRDVPLETOFMLEIESKESFTDGENSPIIYTVLP 123
Db 61 GVLGELRFMCCEFRKLMWTOQRMGTGGRDVPLETOFMLEIESKESFTDGENSPIIYTVLP 123
QY 121 LLEGFRVAVLQNDKNEIEICLSEGDNAVEEDOGILHMYMAGTNPFEVINOAVKAKEXH 173
Db 121 LLEGFRVAVLQNDKNEIEICLSEGDNAVEEDOGILHMYMAGTNPFEVINOAVKAKEXH 173
QY 174 VKAVERKMQTFELHREKKRLPSFVDMGCTWDAFYTDVTAEGVDGLSLSGGPRPRLIIDGMO 233
Db 174 VKAVERKMQTFELHREKKRLPSFVDMGCTWDAFYTDVTAEGVDGLSLSGGPRPRLIIDGMO 233
QY 234 IIDDGMOIENKAKDTECLVDEGAFATRLTGIEKENTFKOKLONNEDMSGLKHLVHGA 293
Db 234 IIDDGMOIENKAKDTECLVDEGAFATRLTGIEKENTFKOKLONNEDMSGLKHLVHGA 293
QY 294 KOHNKAVNYVYHAIAGVGVKPRATGMEHDTALAYVOSPGVLTGNOPODIYVMSLAVHGLVHP 353
Db 294 KOHNKAVNYVYHAIAGVGVKPRATGMEHDTALAYVOSPGVLTGNOPODIYVMSLAVHGLVHP 353
QY 358 KKEHGVASVYVYHAIAGVGVKPRATGMEHDTALAYVOSPGVLTGNOPODIYVMSLAVHGLVHP 358
Db 358 KKEHGVASVYVYHAIAGVGVKPRATGMEHDTALAYVOSPGVLTGNOPODIYVMSLAVHGLVHP 358

[illegible]

```

0Y 178 -EKHMOTFLHREKRRLRPSGLDMFGMCTMAFAITDVAEBVEGLK-----SL 223
Db 181 LKSQKLTSLCPNEFRMPDLNMFGMCTMDAFTYNATKAVKOGSLNCDLTKPALITLCL 240
0Y 224 SOGGTAPRRELLIDDDWQOJENKRAKDATECLVQEOQAFATRLGIKENTKFOKLONNBOJ 283
Db 241 KAGGVATPKFVILIDDGWOVS-GMDETSVEFENMANMANFANRLHIKENHKFOQDGEGHRY 299
0Y 284 S----GLKLHVGAKOHENNVKVVYWHALAGYWGCVKPAATGMEHYDTALAYVPOSPVL 339
Db 300 DDPSTSLGHVIDIKSNNSLKYVYWHALITGYWGCVKPGVSGMEHYESKVAAPVSPGYM 359
0Y 340 GNOPDVLMSLAVHGLVLPKKVFNENYELIAYLASGVDSVKYDVONITETLAGAGGG 399
Db 360 SSENCCLESITKNGILGVNPEKVSFYNDHLSYLASGVDSVKYDVONITETLAGAGGG 419
0Y 400 RVSLSRSHYHAALEASIASNFTDNGCIACWCNHTDLYLSAKOTAIRASDDFPRDPASHT 459
Db 420 RYKLAKKHYQALAEASISRNFDPNGIISCSNHTDLYLSAKKATVIRASDDFPRDPASHT 479
0Y 460 HISSVYANSFLGGERMOPDMDMFSLHPADYHAARAIGGCPIYSDKPGNNHFDLLK 519
Db 480 IHASVAYNTLELGEFMOPODMDMFSLHPMAEYHAARAAGVCAIYVSDKPGQHDENLRL 539
0Y 520 KLVLDGSLVLRQOLGRRPRDLSLFVDPARDKRTSLKINLNKCSGVGVFNOCGAGMCKI 579
Db 540 KVLVDGSLTLRKAKLGRPLR----- 559
0Y 580 EKKTRIDHTSPETLTSYACASDVDLITOVAGAEMLGDFTIYVAYRSGEYIRLPKGVSIPT 639
Db 560 -----ELVYLPKDTSLPT 573
0Y 640 LKVLFEFLPHRCPIQELIAPSISFAAIGLDMFNTOGAVEOYEIHNRAATKTIALSRGRG 699
Db 574 LMPREYEVFTVPVYKEFSDGSKFAVGLMEMNPGCALVSLRYDDEGFKFVYBKMLRSG 633
0Y 700 RGVYSS--QRLPKCYVGAETDF 721
Db 634 LVGVYSSVRRPSRYTVDSDDVEY 656

```

```

RESULT      5
C85025
hypothetical protein AT4g01970 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: C85025
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory Press, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: C85025
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1,807 <STC>
A:Cross-references: GB:MC_001268; NID:g7268581; PIDN:CAB80690.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g01970
A:Map position: 4

```

[illegible]

```

C:Genetics:
A:Gene: AT4g01970
A:Map position: 4

Query Match      32.7%; Score 1307; DB 2; Length 807;
Best Local Similarity 37.3%; Pred. No. 1,2e-92;
Matches 290; Conservative 131; Mismatches 287; Indels 70; Gaps 18;

QY      8  SYVDGKLIVHGKT-IITGVPDNNVLTPGSGGLVY-----GAFVGATAS 50
      : : : : : | | | | | | | | | | | | | | | | | | | | | |
Db      35  NLESGSLCAKADSPILFDVPQNVTFPTSPSSHSISIDAPLPIILRVQNAKGGFLGTKE 94
      : : : : : | | | | | | | | | | | | | | | | | | | | | |

QY      51  HSSSLAVFPWGVLGELRPMCCRRFLMMNMTQRMGTGCGVDVLETFQMLIESKSEETDGEN 110
      : : : : : | | | | | | | | | | | | | | | | | | | | | |
Db      95  SPSDRLTNLSIGREDEDFLSLRFPMWMTAMIKSGSGDQLQAEQWVMLKIPIDS---- 150
      : : : : : | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 111 SPIYTVLPLEGGORAVLQGNDKNEIEICSEJSGNMNVEJDQGLHMYHAGNPEFVI 170
Db 151 ----YVAIIPTLIGAFASALTPEQKGNVTLCSBGSTVKKESSEFSXIAVTHICDNPYLM 206
QY 171 NOAKVAEKHMQFELHREKRRKLRPSCIDMFGWCOTWDAFYDYVAEEVGLSLSGGTP 230
Db 207 KEAFSAKRYAHMNFLLKEKKLPIYDKRGMCTWACVLTYPATIMVGKVEEDGVC 266
QY 231 RFLIIDDGWOI-----ENKADATECLVOBGAFRLTLGIKENTK---OKLONNEOM 283
Db 267 KEVLIIDGMOISINFDGDELDKDA-ENMVLVGEDMTARLISFECKKFRNYKEESIGSDV 325
QY 284 SG-----LKHLVHGAOKOHNVKNVYVHMLAGVMGCVPRATGSEHDTALAYVPOSG 337
Db 326 SGGSMALFTDL---KLRFSLDDIITYVHRLGCMNGVPER--MMDLKAYAPFELSPS 380
QY 338 VLGNQDPIVNDLSLAVHSLGLVHPRKAVFNFYNELHAYLASGVDYKVYDNIETLGAH 397
Db 381 LGATMADLADVKVEAGIIGVHPSKHAHEFDSMHSYLASVGTGAKIDVFTLSEABEH 440
QY 398 GGRSLTSLRSHHMLHSLANFTDNGCICMCMNTDGLSA-KORALVYASDDFPRP- 455
Db 441 GGRVELKAKAYYDGLTESMINFNGTDVILASMOOCNEFFLAKOJSLIGVGDPEWQDPY 500
QY 456 -----ASHRTISVAYNSLFLGERMOPDMOMFSLHPADYHAANAATIGCPIY 505
Db 501 GDPOGYWLGCVHMHICS---YNSIMWGQIOPDMOMFQSDHYCAEYHAASRAICGPHY 557
QY 506 VSDKPG--NENFDLLKLVLPDGSVYLRADLGRPTDLSFYVDPARDRTSLKIMNLNCS 563
Db 558 LSDHLGAKSHNFDLIKRLAFLFDDITPCRYVALPTBDSLEKPNLEFKESILKIFENNFG 617
QY 564 GWVGVCNCGAGMCKIKKTRIRHDTSPGLTTLASVCASDV--DLITQVAAE--WLGDTIV 619
Db 618 GYITGFCNCGAGMPEDEHRRKGYEECTYYSGVVHHYSDIEMQONPAASQYTYGDIYV 677
QY 620 YAIRSEVIRL-PRGVSIPVTLKYLEFELHPCPIOEIAPL-ISAFAILDMENTGAV 677
Db 678 KYOOSEELIFMNSKSEAMKITLESAPDLISFPVPEIVLSSGVRAPDLIMMFCVTV 737
QY 678 FOVELHNAAKTIALTASVRCRGVGYSSORPLKICVYVGAENFDVNFSEGLTFPSIP 735
Db 738 ODMK---TGNSNLRVDYKGGGRMAVSSAPVKCYLINDKEAFKEEETGALSFVP 792

```

RESULT 6
 T09530
 probable seed inhibition protein - chickpea (fragment)
 C:Species: Cicer arietinum (chickpea, garbanzo)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
 C:Accession: T09530
 R:Cervantes, E.
 submitted to the EMBL data Library, February 1996
 A:Reference number: Z16718
 A:Accession: T09530
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-357 <CER>
 A:Cross-references: EMBL:X95875; NID:Q1212811
 A:Experimental source: germinating seed
 C:Genetics:
 A:Gene: slp

Query Match	Similarity	Score	DB	Length
Best Local	68.0%	Pred. No. 46-87		
Matches	247	Conservative	24	Mismatches 75; Indels 17; Gaps 6
Qy	172	QAVKAVKEMQVTLHNEKKRIPSCILDMGCMCTADATYDTYTAGVDEGKSLSSGGTPPR	231	
Db	4	QAVKAVKEMQVTLHNEKKRIPSCILDMGCMCTADATYDTYTAGVDEGKSLSSGGTPPR	63	
Qy	232	FLIIDGWOQIENKAKDATECLVQSGAOFATRLTG--IKENTKFOKKLONNEDMSGILKH	289	

```

Db      64 FLIMTMWGNRLKAVQSGSGCVYQEGQL---LLGDMIRANSKNNKNGNDQIPBLKHL 120
QY      290 VHGAKOHNNHKNVYVNHALAGWGVKPAATGMEHYDTALAYVQSGYLGNOPIVDMS 349
Db      121 VDGAKHHNVKDYVVMHALLAGWGVKPAATGMEHYDTALATSAVTRST-WNDPDIYDMS 179
QY      350 LAVHGLGLVHKKKYFNFYNELHATLASGVDGKVDVQNIITELLAGHGGRVLTGRSYNH 409
Db      180 LAVHGLRLVYHRG-FTSTNETHAYLASGVDGKVDVQITLLRPLVLD-----TVYSRLH 233
QY      410 A-----LEASIASNFTNGCJACMCNHTDGLYSAKQTAIVRASDEYPRDPASHTIHSS 464
Db      234 AAILMRLRLPLLVYLLNNGCJACMCNHTDGLYSAKQTAIVRASDEYPRDPASHTIHSS 293
QY      465 VAYNSLFLGEEMQDPMDMFHSLHPLADYHAAARLIGSCPTIYVDKRPNNHFDLLKTLVLP 524
Db      294 VAYNSLFLGEEMQDPMDMFHSLHPLAAYHAAARAIGSCQPMASYISQATYITLILFLSLVLA 353
QY      525 DGS 527
Db      354 DGS 356

```

RESULT 7
T01717
Hypothetical protein A_IG002N01.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01717
R:Schneel, P.; Magg, L.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana IG002N01.
A:Reference number: Z14407
A:Accession: T01717
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-371 <SCH>
A:CROSS-references: EMBL:AF007269; NID:g2191126; PID:g2191156
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 5/2; 25/3; 103/3; 126/1; 174/2; 240/1; 295/1; 333/2; 349/2
A:Note: A_IG002N01.5

[illegible]

QY 355 LGLVHPKRVENYNELHNLAVLASCSDGVKVDVQNIIETLGAGHGGRVSLTRSYTHALDEAS 414
| | | | |
Db 258 L-----GGGVLAKKKYHOALEAS 275

QY 415 IASNFTDNGCIACMCNHTDGLYSAKOTAIVRASDDFYPRDPASHITIHSSVAYNLSFLGE 474
| : : : : :
Db 276 ISRNPPAG-----KHHDLGYCSHS---OKROLFWHRHNDPASRHTIHASVAYNTLFLEGE 325

OY 475 FMQPMDMFHSLHPADYHAARALIGCGPIYVSKDPGNHNPELLKKLYLPSSYL 529
||||||| : | : ||||| |||||
Db 326 FMPQPMWDISSL-----SWRMCHLCQVIISDKPGQHDFMLKRLVLDGSIL 371

RESULT 8

D90496
hypothetical protein SS03127 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: D90496

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: D90496

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1..649 <KUR>

A:Cross-references: GB:AE006641; NID:g13816550; PIDN:AAK43227.1; GSPDB:GN00155

A:Genetics:

:Gene: SS03127

```

Db      480  PIYIDRHRERINIELRNAVLPNGEIVIRVDEPALITDLEFKDPLRERV-LKLKGRVK 538
Qy      562  CSGVGCVEN 570
          :  ||
Db      539  GYNIAFFN 547

RESULT  9
B11367
probable rep helicase, single-stranded DNA-dependent APNase (rep) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: B11367
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
erson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A11250; MUID:98332770
A:Accession: B11367
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-657 <COL>
A:Cross-references: GB:AE001194; GB:AE000520; NID:93322355; PIDN:AAC65097.1; PID:93322
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0102
C:Superfamily: helicase II

```

[illegible]

RESULT 10

nitrate reductase (NADPH) (EC 1.6.6.3) - Neurospora crassa

516292

C:Species: Neurospora crassa

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S16292; S07176; S34796; S37298

R:Okamoto, P.M.; Fu, Y.H.; Marzluff, G.A.

Mol. Gen. Genet. 227, 213-223, 1991

A:Title: Nit-3, the structural gene of nitrate reductase in Neurospora crassa: nucleotide

A:Reference number: S16292; MUID:91287699

A:Accession: S16292

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1..982 <MOL>

A>Note: the authors translated the codon CCG for residue 140 as Pro

R:Le, K.H.D.; Lederer, F.

EMBO J. 2, 1909-1914, 1983

A:Title: On the presence of a heme-binding domain homologous to cytochrome b5 in Neurosp

A:Reference number: S07176

A:Accession: S07176

A:Molecule type: protein

A:Residues: 'D', 622-623, 'Y', 624, 'IK', 627, 'Y', 628, 'XNXXXT', 635, 'LILHYKX', 643, 'DL', 646, 'K'

C:Genetics:

A:Gene: nit-3

A:Introns: 675/1

C:Complex: homodimer

C:Superfamily: nitrate reductase (NADH); cytochrome b5 core homology; cytochrome-b5 red

C:Keywords: chromoprotein; electron transfer; FAD; flavoprotein; heme; homodimer; iron;

F:121-548/Domain: molybdopterin-binding domain homology <PCO>

F:617-651/Domain: cytochrome b5 core homology <CB5>

F:728-982/Domain: cytochrome-b5 reductase homology <CBR>

F:240/Binding site: molybdopterin (Cys) (covalent) #status predicted

F:499/Disulfide bonds: interchain #status predicted

F:652,675/Binding site: heme iron (His) (axial ligands) #status predicted

F:796,956/Binding site: NADP (Lys, Cys) #status predicted

F:835/Binding site: FAD (Tyr) #status predicted

Query Match	3.1%	Score 125;	DB 1;	Length 982;
Best Local Similarity	22.4%;	Pred. No. 0.41;		
Matches 106;	Conservative 46;	Mismatches 178;	Indels 144;	Gaps 26;
OY 194 SCLDFMGMCTWPAFTYDTVAEGVEGLKSLSGGTPRPFLIIDGMOQIENKAKDA----				249
DB 496 SC---FCMCFMD-----LEIRPLSEL-----				522
OY 250 -TECLVDSGAOPATRLTGIKENTKFOKKLONNEOMSLKH-----LVHGAQOH 297				
DB 523 MDESLALPKEKMYSVLGMNNMFRVYLHNEGDLREPHQPMILSDGMWDRAKKEG 582				
OY 298 NTKKNYVHVALGNGYGVAPATGMENHDYDALAYVPSGVGNQPD--IYDLSLAVHGL 355				
DB 583 NLAN-----GFWGEKVPGA-----EENVVAKEEPVKELSMDEKVTRLITILELRQHD- 629				
OY 356 GLVHEPKKYFN--FYVELHAYILASCGVDGVKVDVONIIETLAGH--GGRVSLRSYHNALE 412				
DB 630 GEEPEWFEVYNGGVYN-----GTFPLE-----GHFGAASITGAAGQDVT 668				
OY 413 ASIASNFDNGCIACMCNHTDGLYSAKQTAIYRASDDPFRPRA----SHTIHISVAYN 468				
DB 669 DEFLIHSNENAMMPETIIGTLTPSAPAAIKSSSS---DPLASDPSRPFILDSKTWN 724				
OY 469 SLEFL--GEEMODPMWFHSL--HPADADYHAARAIGGCPPIYSDKPGNHPDLKLKLYP 524				
DB 725 SAILTFKESVSPDTKIFHIALSHPAQSI-----GLPV-----GQILMRLDPAKP 770				
OY 525 DGSVLAQALP---GRETRDSL-----FVDPARD-----RTSLKIMNLNKC---SGVY 566				
DB 771 TESIIRAYPIPSDGLTERGTLRLVLYIYASPTEDIKGOMQALDALALGKAVEFGKV 830				
OY 567 GVPNCGAGACWCI---EKKTRIHDSPTGLTASVC-ASVDLITLVAGAEWGLD 616				

[illegible]

Db 709 VGCYKGDGFKIFRSENFILRLVTNRRF 736

RESULT 12

S02392

Alpha-2-macroglobulin receptor precursor - human

N:Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor

C:Species: Homo sapiens (man)

C:Date: 14-Aug-1998 #sequence-revision 14-Aug-1998 #text-change 22-Jun-1999

C:Accession: S02392; S30027; I37998; A39210; S12538

R:Herz, J.; Hamann, U.; Rogne, S.; Myklebost, O.; Gausepohl, H.; Stanley, K.K.

EMBO J. 7, 4119-4127, 1988

A:Title: Surface location and high affinity for calcium of a 500-kD liver membrane protein

A:Reference number: S02392; MUID:89210795

A:Accession: S02392

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-4544 <HER>

A:Cross-references: EMBL:X13916; NID:g34338; PIDN:CAA32112.1; PID:g34339

R:Kristensen, T.

submitted to the EMBL Data Library, October 1990

A:Reference number: S30027

A:Accession: S30027

A:Molecule type: mRNA

A:Residues: 3275-3864 <KRI>

A:Cross-references: EMBL:X55077

R:Herz, J.; Kowal, R.C.; Goldstein, J.L.; Brown, M.S.

EMBO J. 9, 1769-1776, 1990

A:Title: Proteolytic processing of the 600 kd low density lipoprotein receptor-related protein

A:Reference number: S12538; MUID:90269210

A:Contents: annotation; site of proteolytic cleavage

R:Kuit, H.; Herz, J.; Stanley, K.K.

Biochim. Biophys. Acta 1009, 229-236, 1989

A:Title: Structure of the low-density lipoprotein receptor-related protein (LRP) promoter

A:Reference number: I37998; MUID:90089395

A:Accession: I37998

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-11 <RES>

A:Cross-references: EMBL:X15424; NID:g34408; PIDN:CAA3464.1; PID:g34409

R:Strickland, D.K.; Ashcom, J.D.; Williams, S.; Burgess, W.H.; Migliorini, M.; Argaves, J.

Biol. Chem. 265, 17401-17404, 1990

A:Title: Sequence identity between the alpha2-macroglobulin receptor and low density lipoprotein receptor

A:Reference number: A39210; MUID:91009181

A:Accession: A39210

A:Status: preliminary

A:Molecule type: protein

A:Residues: 150-166;234-238, 'X', 240-245, 'X', 247-252, 'G', 686-695;902-916;1096-1109; 'S', 17

C:Genetics: GDB:LRP1; APR; LRP; A2MR

A:Cross-references: GDB:119694; OMIM:107770

A:Map position: 12q13.1-12q13.3

C:Complex: The alpha2-macroglobulin receptor complex consists of noncovalently-associated protein (see PIR:A396875).

C:Superfamily: Alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding C:Keywords: beta-hydroxyaspartate; beta-hydroxyaspartic acid; calcium binding; glycoprotein

F:1-19/Domain: signal sequence #status predicted <SIG>

F:207-3943/Product: alpha-2-macroglobulin receptor 515k chain #status predicted <S15k>

F:27-64/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:72-106/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:115-148/Domain: EGF homology <EG1>

F:154-188/Domain: EGF homology <EG2>

F:198-239/Domain: LDL receptor WYTD-containing repeat homology <YW01>

F:240-281/Domain: LDL receptor WYTD-containing repeat homology <YW02>

F:292-334/Domain: LDL receptor WYTD-containing repeat homology <YW03>

F:335-378/Domain: LDL receptor WYTD-containing repeat homology <YW04>

F:379-420/Domain: LDL receptor WYTD-containing repeat homology <YW05>

F:427-468/Domain: LDL receptor WYTD-containing repeat homology <YW06>

F:478-519/Domain: EGF homology <EG3>

F:517-613/Domain: LDL receptor WYTD-containing repeat homology <YW07>

F:614-659/Domain: LDL receptor WYTD-containing repeat homology <YW08>

F:660-710/Domain: LDL receptor WYTD-containing repeat homology <YW09>

F:711-752/Domain: LDL receptor WYTD-containing repeat homology <YW10>

F:753-799/Domain: LDL receptor WYTD-containing repeat homology <YW11>

F:807-842/Domain: EGF homology <EG4>

F:854-890/Domain: LDL receptor ligand-binding repeat homology <LDL3>

F:895-931/Domain: LDL receptor ligand-binding repeat homology <LDL4>

F:936-971/Domain: LDL receptor ligand-binding repeat homology <LDL5>

F:976-1011/Domain: LDL receptor ligand-binding repeat homology <LDL6>

F:1015-1051/Domain: LDL receptor ligand-binding repeat homology <LDL7>

F:1062-1097/Domain: LDL receptor ligand-binding repeat homology <LDL8>

F:1104-1140/Domain: LDL receptor ligand-binding repeat homology <LDL9>

F:1145-1182/Domain: LDL receptor ligand-binding repeat homology <LDL10>

F:1185-1221/Domain: EGF homology <EG5>

F:1227-1261/Domain: EGF homology <EG6>

F:1269-1308/Domain: LDL receptor WYTD-containing repeat homology <YW12>

F:1309-1355/Domain: LDL receptor WYTD-containing repeat homology <YW13>

F:1356-1398/Domain: LDL receptor WYTD-containing repeat homology <YW14>

F:1399-1445/Domain: LDL receptor WYTD-containing repeat homology <YW15>

F:1446-1488/Domain: LDL receptor WYTD-containing repeat homology <YW16>

F:1489-1531/Domain: LDL receptor WYTD-containing repeat homology <YW17>

F:1540-1578/Domain: EGF homology <EG7>

F:1583-1626/Domain: LDL receptor WYTD-containing repeat homology <YW18>

F:1627-1669/Domain: LDL receptor WYTD-containing repeat homology <YW19>

F:1670-1713/Domain: LDL receptor WYTD-containing repeat homology <YW20>

F:1714-1753/Domain: LDL receptor WYTD-containing repeat homology <YW21>

F:1754-1796/Domain: LDL receptor WYTD-containing repeat homology <YW22>

F:1797-1846/Domain: LDL receptor WYTD-containing repeat homology <YW23>

F:1850-1886/Domain: EGF homology <EG8>

F:1934-1976/Domain: LDL receptor WYTD-containing repeat homology <YW24>

F:1977-2019/Domain: LDL receptor WYTD-containing repeat homology <YW25>

F:2020-2063/Domain: LDL receptor WYTD-containing repeat homology <YW26>

F:2064-2105/Domain: LDL receptor WYTD-containing repeat homology <YW27>

F:2106-2151/Domain: EGF homology <EG9>

F:2159-2194/Domain: LDL receptor WYTD-containing repeat homology <YW28>

F:2199-2241/Domain: LDL receptor WYTD-containing repeat homology <YW29>

F:2253-2294/Domain: LDL receptor WYTD-containing repeat homology <YW30>

F:2344-2388/Domain: LDL receptor WYTD-containing repeat homology <YW31>

F:2389-2429/Domain: LDL receptor WYTD-containing repeat homology <YW32>

F:2430-2473/Domain: LDL receptor WYTD-containing repeat homology <YW33>

F:2482-2511/Domain: EGF homology <EG10>

F:2524-2561/Domain: LDL receptor ligand-binding repeat homology <LDL11>

F:2566-2600/Domain: LDL receptor ligand-binding repeat homology <LDL12>

F:2605-2639/Domain: LDL receptor ligand-binding repeat homology <LDL13>

F:2652-2688/Domain: LDL receptor ligand-binding repeat homology <LDL14>

F:2696-2739/Domain: LDL receptor ligand-binding repeat homology <LDL15>

F:2734-2769/Domain: LDL receptor ligand-binding repeat homology <LDL16>

F:2774-2812/Domain: LDL receptor ligand-binding repeat homology <LDL17>

F:2818-2853/Domain: LDL receptor ligand-binding repeat homology <LDL18>

F:2858-2897/Domain: LDL receptor ligand-binding repeat homology <LDL19>

F:2904-2939/Domain: LDL receptor ligand-binding repeat homology <LDL20>

F:2944-2980/Domain: EGF homology <EG11>

F:2986-3021/Domain: EGF homology <EG12>

F:3029-3068/Domain: LDL receptor WYTD-containing repeat homology <YW34>

F:3069-3113/Domain: LDL receptor WYTD-containing repeat homology <YW35>

F:3114-3156/Domain: LDL receptor WYTD-containing repeat homology <YW36>

F:3157-3200/Domain: LDL receptor WYTD-containing repeat homology <YW37>

F:3201-3241/Domain: LDL receptor WYTD-containing repeat homology <YW38>

F:3242-3284/Domain: LDL receptor WYTD-containing repeat homology <YW39>

F:3294-3330/Domain: EGF homology <EG13>

F:3334-3369/Domain: LDL receptor ligand-binding repeat homology <LDL21>

F:3374-3408/Domain: LDL receptor ligand-binding repeat homology <LDL22>

F:3413-3448/Domain: LDL receptor ligand-binding repeat homology <LDL23>

F:3454-3489/Domain: LDL receptor ligand-binding repeat homology <LDL24>

F:3493-3531/Domain: LDL receptor ligand-binding repeat homology <LDL25>

F:3536-3570/Domain: LDL receptor ligand-binding repeat homology <LDL26>

F:3575-3609/Domain: LDL receptor ligand-binding repeat homology <LDL27>

F:3613-3647/Domain: LDL receptor ligand-binding repeat homology <LDL28>

F:3654-3690/Domain: LDL receptor ligand-binding repeat homology <LDL29>

F:3695-3731/Domain: LDL receptor ligand-binding repeat homology <LDL30>

F:3741-3776/Domain: LDL receptor ligand-binding repeat homology <LDL31>

F:3785-3822/Domain: EGF homology <EG14>

F:3828-3860/Domain: EGF homology <EG15>

F:3868-3911/Domain: LDL receptor WYTD-containing repeat homology <YW40>

F:3912-3969/Domain: LDL receptor WYTD-containing repeat homology <YW41>

F:3944-4544/Product: alpha-2-macroglobulin receptor 85k chain #status predicted <85k>
 F:3944-4420/Domains: 85k chain extracellular #status predicted <Ext>
 F:3970-4012/Domains: LDL receptor WTRD-containing repeat homology <YW42>
 F:4013-4056/Domains: LDL receptor WTRD-containing repeat homology <YW43>
 F:4057-4099/Domains: LDL receptor WTRD-containing repeat homology <YW44>
 F:4100-4142/Domains: LDL receptor WTRD-containing repeat homology <YW45>
 F:4151-4183/Domains: EGF homology <EG16>
 F:4200-4231/Domains: EGF homology <EG17>
 F:4236-4267/Domains: EGF homology <EG18>
 F:4272-4303/Domains: EGF homology <EG19>
 F:4308-4339/Domains: EGF homology <EG20>
 F:4344-4374/Domains: EGF homology <EG21>
 F:4377-4408/Domains: EGF homology <EG22>
 F:4421-4444/Domains: transmembrane #status predicted <TM>
 F:4445-4544/Domains: Intracellular #status predicted <INT>
 F:166-2958/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
 F:2958/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
 F:4075-4125,4278/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.9%; Score 117.5; DB 1; Length 4544;
 Best Local Similarity 20.8%; Pred. No. 16; Mismatches 178; Indels 145; Gaps 25;
 Matches 98; Conservative 51;

OY 71 CERFLMMWTOGMGTC-----GRDVPLETOFMLESKESEPTDENSEPIITYTLPL 121
 DB 3828 CLRF-----GTCSOLCENNTKGGHLCSCARFM---KTHNCKMGSE--YQVLXIA 3873
 OY 122 LEGOFAVLQGNKNEIEICLESQNAVEIT-QGLHM---VY--MHAGTNPPEVINO 173
 DB 3874 DNEIRSLFPGHPHSAEQAFQ-GDESVRIDAMDVHAKGVWYTMWTGTITISYRSLPPA 3932
 OY 174 VKAV--EKHM-----FELHREKKRLPSCLDWFGWCWTFADFYDVAEGVE-----EG 219
 DB 3933 ADPTTSNRRRIODRGVHNLISGLKMPRGIA-IDWAGANVYWDSCGADVLEVYQMKGEN 3991
 OY 220 LKSLSGG-TTPRFLIID-----DGM---OQIENKAKAT--ECLVQSGAFATRLT 265
 DB 3992 KRTILSGMIDEPHALVVDPLRGTMWSDMGNHPKLETAAMQGLRETLVQONIQMPTGL- 4050
 OY 266 GIKENTKFOKKLQNNQNSGKLHLVHGAKOHNNKANYVHNAIACGYGCVAPRATGMHY 325
 DB 4051 -----AVDYHNERL-----YMDALSLYIG----- 4070
 OY 326 DFLALVPQSPGVLPNOPIVMDSLAVHGLVHPKVFNFYNELHAIASCGVDGYVD 385
 DB 4071 -----SIRLNGTDPVADSKR-----GLSHFPSIDVEDYIY-----GVYI 4108
 OY 386 VONITETLACGHGRVSLTRSYHALLASI-----ASNETDNG-----CI-----A 426
 DB 4109 NNRVFKIRHFGHSPVNLGLSHASDVLYHQHKOPEVTNCDKRCCEMLCLLSPSGPV 4168
 OY 427 CMCHNTDGLYSKQTAIVASDPFPPROPASHTIHSSVAANSLPLGEMOP 478
 DB 4169 CTCPCMKRILDNG--TCVPVSPPTPPDAPRGTCLMDCPENGSGCLNARROP 4218

RESULT 13
 G86546
 polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (strat

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 15-Jun-2001
 C:Accession: G86546

Nucleic Acids Res. 28, 2311-2314, 2000
 A>Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A:Reference number: A86491; MOID:20330349

A:Accession: G86546
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-928 <STO>

A:Cross-References: GB:BA000008; NID:98978821; PIDN:BA98657.1; GSPDB:GM00142
 A:Experimental source: strain J138

C:Genetics:
 A:Gene: pmp_10
 C:Superfamily: Chlamydia pneumoniae polymorphic outer membrane protein G

Query Match 2.9%; Score 115; DB 2; Length 928;
 Best Local Similarity 24.2%; Pred. No. 2.3;
 Matches 107; Conservative 39; Mismatches 145; Indels 152; Gaps 23;

OY 338 VLGNOPDIWMDSLAVHGLVHPKVFNFYNELHAIASCGVDGYKVOVONITETLGAGH 397
 DB 271 ISGNOSVTFSGNOAVANGSAIYAKRL-----TLASGGGGISFS--NNIVGGTTAGN 320
 OY 398 GGRVSLTRSYHALLASIASNFTDNCIACMCHNTDGLYSKQTAIVASDPFPPROPAS 457
 DB 321 GGASITLAGECSLSAE--AGDITFNG-----NAIVATF--PQTKR 358
 OY 458 HTHTISSVAIYNSLFLGEPQPMDFHSLHPRADYHAARAIGGCPRIYSDKPGHNDL 517
 DB 359 NSIDIGSTAKIT-----NLRAISGHSIF----- 381
 OY 518 LKKLVLPDSVLAQLPGRPTRSLFVDP-----ARDTSILKINLNLKC-----SG 564
 DB 382 -----PYDEPTANTADSDTL--NLNKADAGNSTDYSG 413
 OY 565 VVGVFNCQAGMCKIEKTRIHDTs---PGLTAS--VCASDVLD---ITQVAGAEWL 614
 DB 414 SI-VFS--GEKLSDEBAKAVADNLTSLKOPVTLTACNLKRGVTLDTKGFOTAGSSVT 470
 OY 615 GDTIVAYRSGEYTRPKGVSTPYT-----LVLEELHPQPIQELIASISFAIG--- 666
 DB 471 MDAGTTLKASTERTVL-TGLSTIPVDSLGGSKVY-----IASASKNVALSGPL 520
 OY 667 LLDMPNTGAVQEVETIHNRAATKTIALSVGRGREGVSSORPLKCVGGAETDENYDSE 726
 DB 521 LLD--NQNAYENHDLCKQODESFVQLSALGTAT-----TTDPAVAPTV-ATPTHYGYQGT 573
 OY 727 TGLT---TFSDVSPENYRWS 745
 DB 574 WGMWVVDVTASTPTKTATLATMT 596

RESULT 14
 G81591
 polymorphic outer membrane protein G family CP0303 [imported] - Chlamydia pneumoniae (s

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 15-Jun-2001
 C:Accession: G81591

R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 , C.; Dodson, R.; Gilm, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salze
 , C.; Nucleic Acids Res. 28, 1397-1406, 2000
 A>Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MOID:20150255

A:Accession: G81591
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-928 <RBA>

A:Cross-References: GB:AE002192; GB:AE002161; NID:97189226; PIDN:AAF38160.1; PID:9718

A:Experimental source: strain AR39, HL cells
 C:Genetics:
 C:Superfamily: Chlamydia pneumoniae polymorphic outer membrane protein G

Query Match 2.9%; Score 115; DB 2; Length 928;
 Best Local Similarity 24.2%; Pred. No. 2.3;
 Matches 107; Conservative 39; Mismatches 145; Indels 152; Gaps 23;

OY 338 VLGNOPDIWMDSLAVHGLVHPKVFNFYNELHAIASCGVDGYKVOVONITETLGAGH 397
 DB 271 ISGNOSVTFSGNOAVANGSAIYAKRL-----TLASGGGGISFS--NNIVGGTTAGN 320
 OY 398 GGRVSLTRSYHALLASIASNFTDNCIACMCHNTDGLYSKQTAIVASDPFPPROPAS 457


```

Db      321  GGAISILAGECSLSAE-ADDITFNG-----NAIVATT-----POTTKR 358
Qy      458  HTIISSVAANSFLGFGKOPDMDMESHLPADYHAARAIGCPITYSDKGNHNFDL 517
Db      359  NSIDIGSTAKIT-----NRAISGHSIF-----381
Qy      518  LKRLVLDGSVLRKQLRGRTROSLFVDP-----ARDRTSLKIKWLNKC-----SG 564
Db      382  -----FYDPTTANTADSTDTL-----NLRNADAGNSTDYSG 413
Qy      565  VGVFNCOGAGMCKIEKTRIHDT-----PGTLTAS--VCASDVDP-----ITVOGAEML 614
Db      414  SI-VFS--GEKLSDEDAKVADNLTSLKQPVTLTACNMLVLRGVTLDTGTGFOTASSVY 470
Qy      615  GDTIVVARGGEYIRLPKGVSIYV-----LKVLEFLHFPCIOIAPISISFAIG-- 666
Db      471  MDGTTILKASTEEVTL-TGISIPVDSLGEKKVY-----IAASAASKNVALSGPIL 520
Qy      667  LIDMFNTGAVEQVEIHNRATKTIASVGRGFRGVSSQRPKCVGGAETDEKDYSE 726
Db      521  LLD--NOGNAYENHDJGKODFSFVQSLALGTAT-----TTDVPAVPTV-ATPHHGYQGT 573
Qy      727  TGLT-----TFSTIPVSPEMYRMS 745
Db      574  WGMTWVDTDASTPKTKTATLAWT 596

```

RESULT 15

G86474

probable protein gypsy/Ty3 element polyprotein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: G86474

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.;
ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

A:Authors: Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shlum, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: G86474

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1447 <STO>

A:Cross-references: GB:AE005172; NID:g11034946; PIDN:AAG27103.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

```

Query Match          2.9%; Score 114.5; DB 2; Length 1447;
Best Local Similarity 20.2%; Pred. No. 4.8;
Matches 86; Conservative 58; Mismatches 142; Indels 139; Gaps 21;

Qy      110  NSPIITYVLLPLLEGQFAVLQGNDKNEIEICESGNAVETD---OGLHWYMHAGTNP 166
Db      854  NAPVL---ALPYEDKQF-----MVEDACGGGIRAVILMOKG-HP 888
Qy      167  FEVINQAVKAEKHMOTFLHREKKRLP---SCLDMFGWCTWDATYTDYTAGS---VEEG 219
Db      889  LAYISROLKGKOLHSIY---EKELLAFIFAVRKWRHYLPSHFITKTDQRSIKYLLDQR 945
Qy      220  LKSLSGGTTPREFLIIDSGWOQIENKADATECLVQ--EGAQFATRLTGIKENTKFOKKLQ 278
Db      946  LNTPYOQOQMLPRLFEDEIYRQCKENLVADALSRYEGSEVLHMAI.SIIVE-CDFLKEIQ 1004
Qy      279  NNEQMSG-LKHLVHGAKOHNNVKNYVNHALAGYGVGKVPATGMEHNDYDLAIPVQSPG 337
Db      1005  VAYESDGVLKDLISALQOHPRAKKHYSW-----SQD 1035

```

```

Qy      338  VLGNQPDIVYMS-----LAVHGLG-----LVHPKKVFENFY-----NELHAYIA 375
Db      1036  ILRRRSKIIVPADVETLTKLQWLHCSCMGGRSGRSDASHQVYKSLFYKKGVAKDIOAFIR 1095
Qy      376  SCGV-----DGVKVDY-QNIETLQAGHGRV-----SLFRSY 407
Db      1096  SCGTCQOCKSDNAAYPGILOPLPIPKIWCDSYMDFIEGLPNSGGKSVIMVVDRLSKAA 1155
Qy      408  H-----HALEA-SIASNFTDN-----GCIACMCHNTDGLYSAK-----QTAIVRASD 448
Db      1156  HTVALAHFYSALTVAQAFLDNVYKHHGCPSTISVDVLTFSDFWKEFFKLLQGVLELMS 1215
Qy      449  DEYPR 453
Db      1216  AYHPQ 1220

```

Search completed: June 4, 2002, 09:06:42
Job time: 220 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 4, 2002, 09:06:13 ; Search time 13.31 Seconds

(without alignments)
2181.793 Million cell updates/sec

Title: US-09-810-186-1
Perfect score: 3996
Sequence: 1 MVTMPKISVNGKLVHGT.....TFSTPSPPEMYRMSIELQV 750

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
- score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125	3.1	982	1	NIA_NEUCR
2	117.5	2.9	4544	1	LRL1_HUMAN
3	115	2.9	928	1	PM10_CHLPN
4	109.5	2.7	1338	1	VGR1_HUMAN
5	108	2.7	714	1	PBPF_BACSU
6	106	2.7	976	1	PMEI_CHLMU
7	105.5	2.6	555	1	PTB_RAT
8	103.5	2.6	587	1	PICP_PSESR
9	103.5	2.6	2210	1	RRLP_EROSM
10	102	2.6	2206	1	POLG_POL2L
11	102	2.6	2207	1	POLG_POL2L
12	101.5	2.5	1430	1	RPOC_XYLFA
13	101	2.5	2208	1	POLH_POL1M
14	101	2.5	4466	1	DYHC_ANTCR
15	101	2.5	4466	1	DYHC_ANTCR
16	100	2.5	2206	1	POLG_CXA21
17	99.5	2.5	815	1	RGRG_HUMAN
18	99	2.5	2214	1	POLG_CXA24
19	98.5	2.5	525	1	ACU8_NEUCR
20	98.5	2.5	537	1	2029_XENLA
21	98	2.5	776	1	ISOA_PSEYA
22	97.5	2.4	843	1	ARM_DROME
23	97	2.4	894	1	WTRA_BACSU
24	97	2.4	914	1	COPP_DROME
25	97	2.4	1019	1	VP3_RDVA
26	97	2.4	1536	1	GLSF_ANTSP
27	96.5	2.4	4303	1	MS1L_HUMAN
28	96	2.4	424	1	MS1L_HUMAN
29	96	2.4	1403	1	BIRE_MOUSE
30	95	2.4	267	1	EBAG_FLASP
31	95	2.4	880	1	LYTD_BACSU
32	95	2.4	1019	1	VP3_RDVF
33	95	2.4	2205	1	POLG_POL2W

ALIGNMENTS

34	94.5	2.4	353	1	TCT2_PHRPO	061064 physarum po
35	94.5	2.4	610	1	CN42_RAT	p14645 rattus norv
36	94.5	2.4	1816	1	LM44_MOUSE	p97927 mus musculu
37	94	2.4	530	1	KPY1_RAT	p11960 rattus norv
38	94	2.4	551	1	CBX4_MOUSE	055187 mus musculu
39	94	2.4	776	1	ISOA_PSESP	p26501 pseudomonas
40	94	2.4	1112	1	ATB2_OREMO	p8165 oreochromis
41	94	2.4	2415	1	PCBM_HUMAN	p16112 homo sapien
42	94	2.4	4393	1	PCBM_HUMAN	p8160 homo sapien
43	93.5	2.3	551	1	AAT1_BACSU	p53001 bacillus su
44	93.5	2.3	551	1	CH62_SYNY3	p22034 synechocyst
45	93.5	2.3	682	1	OE66_NPVOF	010305 oryza pseu

RESULT 1
ID NIA_NEUCR STANDARD; PRT; 982 AA.
AC P08619;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Nitrate reductase [NADPH] (EC 1.6.6.3) (NR).
GN NIT-3.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=91287699; PubMed=1829499;
RA Okamoto P.M., Fu Y.-H., Marzluf G.A.;
RT "Nit-3, the structural gene of nitrate reductase in Neurospora
crassa: nucleotide sequence and regulation of mRNA synthesis and
turnover.";
RL Mol. Gen. Genet. 227:213-223(1991).
RN [2]
RP PRELIMINARY PARTIAL SEQUENCE AROUND HIS-652.
RA Le K.H.D., Lederer F.;
RT "On the presence of a heme-binding domain homologous to cytochrome b5
in Neurospora crassa assimilatory nitrate reductase.";
RL EMBO J. 2:1909-1914(1983).
RN [3]
RP MUTANTS.
RC STRAIN=74-OR23-1A;
RX MEDLINE=93241176; PubMed=8479443;
RA Okamoto P.M., Garrett R.H., Marzluf G.A.;
RT "Molecular characterization of conventional and new repeat-induced
mutants of nit-3, the structural gene that encodes nitrate reductase
in Neurospora crassa.";
RL Mol. Gen. Genet. 238:81-90(1993).
RN [4]
RP MUTAGENESIS.
RX MEDLINE=93360901; PubMed=8355655;
RA Okamoto P.M., Marzluf G.A.;
RT "Nitrate reductase of Neurospora crassa: the functional role of
individual amino acids in the heme domain as examined by
site-directed mutagenesis.";
RL Mol. Gen. Genet. 240:221-230(1993).
CC -1- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -1- CATALYTIC ACTIVITY: NADPH + nitrate = NADP(+) + nitrite + H(2)O.
CC -1- COFACTOR: EACH SUBUNIT CONTAINS 1 EQUIVALENT OF FAD, HEME IRON,
AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME GROUP IS
CALLED CYTOCHROME B-557.
CC -1- PATHWAY: NITRATE ASSIMILATORY PATHWAY.
CC -1- SUBUNIT: HOMODIMER.
CC -1- INDUCTION: ITS EXPRESSION IS HIGHLY REGULATED AND RESPONDS RAPIDLY
TO NITRATE INDUCTION AND TO NITROGEN REPRESSION.
CC -1- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE

RT		"NMR solution structure of complement-like repeat CR8 from the low
RJ		density lipoprotein receptor-related protein.";
RL	J. Biol. Chem.	274:14130-14136(1999).
CC	-I-	FUNCTION: INVOLVED IN THE PLASMA CLEARANCE OF CHYLOMICRON REMNANTS
CC		AND ACTIVATED ALPHA 2-MACROGLOBULIN, AS WELL AS THE LOCAL
CC		METABOLISM OF COMPLEXES BETWEEN PLASMINOGEN ACTIVATORS AND THEIR
CC		ENDOGENOUS INHIBITORS.
CC	-I-	SUBCELLULAR LOCATION: Type I membrane protein.
CC	-I-	TISSUE SPECIFICITY: MOST ABUNDANT IN LIVER, BRAIN AND LUNG.
CC	-I-	PTM: CLEAVED INTO A 85 kDa MEMBRANE-SPANNING SUBUNIT (LRP-85) AND
CC		A 515 kDa LARGE EXTRACELLULAR DOMAIN (LRP-515) THAT REMAINS NON-
CC		COVALENTLY ASSOCIATED.
CC	-I-	SIMILARITY: CONTAINS 22 EGF-LIKE DOMAINS.
CC	-I-	SIMILARITY: CONTAINS 31 LDL-RECEPTOR CLASS A DOMAINS.
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC		or send an email to license@sdb.sdb.ch).
CC		-----
DR	EMBL:	X13916; CAA32112.1; .
DR	EMBL:	AF058427; AAC64265.1; .
DR	PDB:	1CR8; 06-JAN-99.
DR	MIM:	107770; .
DR	Interpro:	IPR000152; Asx_hydroxyl.
DR	Interpro:	IPR000561; EGF-like.
DR	Interpro:	IPR001881; EGF_Ca.
DR	Interpro:	IPR002172; LDL_recept_A.
DR	Interpro:	IPR000033; ldl_receptor_rep.
DR	-Pfam:	PF00008; EGF_16.
DR	-Pfam:	PF00057; ldl_recept_a; 31.
DR	-Pfam:	PF00058; ldl_recept_b; 33.
DR	PRINTS:	PR00261; LDLRECEPTOR.
DR	SMART:	SM00179; EGF_CA; 3.
DR	SMART:	SM00001; EGF_Like; 19.
DR	SMART:	SM00192; LDla; 31.
DR	SMART:	SM00135; Ly; 32.
DR	PROSITE:	PS00010; ASX_HYDROXYL; 3.
DR	PROSITE:	PS00022; EGF_1; 5.
DR	PROSITE:	PS01186; EGF_2; 8.
DR	PROSITE:	PS01187; EGF_CA; 2.
DR	PROSITE:	PS01209; LDLRA_1; 27.
DR	PROSITE:	PS50068; LDLRA_2; 31.
KW	Receptor:	Transmembrane Repeat; Endocytosis; Glycoprotein;
KW	Signal:	Calcium-binding; EGF-like domain; Coated pits; 3D-structure.
FT	SIGNAL	1
FT		19
FT	CHAIN	20 4544
FT		LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED
FT		PROTEIN 1.
FT	DOMAIN	20 4419
FT	TRANSMEM	4420 4444
FT	DOMAIN	4445 4544
FT	DOMAIN	25 66
FT	DOMAIN	70 110
FT	DOMAIN	111 149
FT	DOMAIN	150 189
FT	DOMAIN	474 520
FT	DOMAIN	803 843
FT	DOMAIN	852 892
FT	DOMAIN	893 933
FT	DOMAIN	934 973
FT	DOMAIN	974 1013
FT	DOMAIN	1013 1053
FT	DOMAIN	1060 1099
FT	DOMAIN	1102 1142
FT	DOMAIN	1143 1182
FT	DOMAIN	1183 1222
FT	DOMAIN	1223 1262
FT	DOMAIN	1262 1579
FT	DOMAIN	1536 1579
FT	DOMAIN	1846 1887
FT	DOMAIN	2155 2195
FT		EGF-LIKE 8.
FT		EGF-LIKE 9.
FT		EGF-LIKE 10.
FT		LDL-RECEPTOR CLASS A 3.
FT		LDL-RECEPTOR CLASS A 4.
FT		LDL-RECEPTOR CLASS A 5.
FT		LDL-RECEPTOR CLASS A 6.
FT		LDL-RECEPTOR CLASS A 7.
FT		LDL-RECEPTOR CLASS A 8.
FT		LDL-RECEPTOR CLASS A 9.
FT		LDL-RECEPTOR CLASS A 10.
FT		EGF-LIKE 5.
FT		EGF-LIKE 6.
FT		EGF-LIKE 7.
FT		EGF-LIKE 8.
FT		EGF-LIKE 9.
FT		EGF-LIKE 10.
FT		EXTRACELLULAR (POTENTIAL).
FT		CYTOSOLASMIC (POTENTIAL).
FT		LDL-RECEPTOR CLASS A 1.
FT		LDL-RECEPTOR CLASS A 2.
FT		EGF-LIKE 1.
FT		EGF-LIKE 2.
FT		EGF-LIKE 3.
FT		EGF-LIKE 4.
FT		LDL-RECEPTOR CLASS A 3.
FT		LDL-RECEPTOR CLASS A 4.
FT		LDL-RECEPTOR CLASS A 5.
FT		LDL-RECEPTOR CLASS A 6.
FT		LDL-RECEPTOR CLASS A 7.
FT		LDL-RECEPTOR CLASS A 8.
FT		LDL-RECEPTOR CLASS A 9.
FT		LDL-RECEPTOR CLASS A 10.
FT		EGF-LIKE 5.
FT		EGF-LIKE 6.
FT		EGF-LIKE 7.
FT		EGF-LIKE 8.
FT		EGF-LIKE 9.
FT		EGF-LIKE 10.
FT		EXTRACELLULAR (POTENTIAL).
FT		CYTOSOLASMIC (POTENTIAL).
FT		LDL-RECEPTOR CLASS A 1.
FT		LDL-RECEPTOR CLASS A 2.
FT		EGF-LIKE 1.
FT		EGF-LIKE 2.
FT		EGF-LIKE 3.
FT		EGF-LIKE 4.
FT		LDL-RECEPTOR CLASS A 3.
FT		LDL-RECEPTOR CLASS A 4.
FT		LDL-RECEPTOR CLASS A 5.
FT		LDL-RECEPTOR CLASS A 6.
FT		LDL-RECEPTOR CLASS A 7.
FT		LDL

Query Match	Similarity	Score	DB 1	Length
Best Local	96	20.88	51	145
Matches	96	Conservative	51	Mismatches 178; Indels 145; Gaps 25
71 CFFRKMMVQRRGTC-----GRVPLEYQMLIESKESSETDGENSPIITYVLP 121				

```

Db 3628 CLRF-----GTCGSQLCNNTKGGHLCSCARNFM---KTHNCKAESE--YQYLTYA 3673
QY 122 LEGOFRAVLQGNDRKNEITEICLSEGSNAVETD-QGLHM-----VY---MHAQTFPEVYINQA 173
Db 3874 DDNEIRSLFPHPHSAVEQAFQ-GDESVRIDAMVHKAGRWWTNMHTGISRSRPPA 3932
QY 174 VKAV--EKHMQ-----TFHREKKRLPSCLDWFGCTWDAFYTDVAEGVE-----EG 219
Db 3933 APPTTSNHRROIIDRGVTHLNIISGLKMPRGIA-IDWVAGVNYWTDSGRDVYEAQMKGEN 3991
QY 220 LKLSLOG-PPRRLIID-----DGW---OOENKAKDAT--ECLVGAQOFATRLT 265
Db 3992 RKTLLSGMIDEPHAIIVDPRLGRTWSDMGNHPKIEETAADGTIRETLVDNIDNITGL- 4050
QY 266 GIKENTKFOKKLONNEQMSGLKHLVHGAKOHNVKNYVYMHALAGVGGVPAATGMEHY 325
Db 4051 -----AVDYHNERL-----YWADAKLSVIG--- 4070
QY 326 DTAALAYPQSPGVGNOPDIYMDSLAHGLGLVHPKVFNFYNELHAYLASCGVAGKYVD 385
Db 4071 -----SIRLNGTDEIVADSKR-----GLSHPFSDVFEEDYI-----GVTYI 4108
QY 386 VONITETLGAHGGRVSLTRSYHHALEASI-----ASNFTDNG-----CI-----A 426
Db 4109 NNRVFKIKRFGHSPVLYNLTGSLSHASDVLYLHOKKQREVTNPDCKRKCCEMCLLSPSGPV 4168
QY 427 CMCNHTDGLYSAKOTAIVRASDDFPRDPASHTHISSAVNSLFLGEMQP 478
Db 4169 CTCPRGKRRLDNG--TCVPVPSPPTPPDAPRPGTCLACFCNGSGCLNARROP 4218

```

```

RESULT 3
PM10_CHLPN STANDARD; PRT: 928 AA.
ID PM10_CHLPN
AC Q9RB65; Q9RB64; Q9S6P2; 086163;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmpl0 precursor (Polymorphic membrane
protein 10) (Outer membrane protein 5).
GN PM10 OR OMP5 OR CP0303.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310;
RX MEDLINE=20007584; PubMed=10539856;
RA Christiansen G., Boesen T., Hjerno K., Dagaard L., Mygind P.,
Madsen A.S., Knudsen K., Falk E., Birkelund S.;
RT "Molecular biology of Chlamydia pneumoniae surface proteins and their
role in immunopathogenicity."
RL Am. Heart J. 138:S491-S495(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey E.K., Peterson J., Ullerbach T., Berry K., Baas S.,
Linder K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia
pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shitai M., Hatakeyama H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
Shiba T., Ishii R., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138

```

```

RT from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [4]
RP SEQUENCE OF 1-914 FROM N.A.
RC STRAIN=CWL-029/VR-1310;
RX MEDLINE=99081766; PubMed=9864239;
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding 97- to 99-kilodalton outer
membrane proteins of Chlamydia pneumoniae."
RL Infect. Immun. 67:375-383(1999).
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC CC
CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
CC EMBL: AJ133034; CAB37071.1; -
CC DR EMBL: AE002192; AAF38160.1; -
CC DR EMBL: AF002546; BAA98657.1; -
CC DR EMBL: AJ001311; CAA04671.1; -
CC DR PHC1-2DPAGE: 086163; -
CC DR TIGR: CP0303; -
CC DR InterPro: IPR003357; OMP.
CC DR Pfam: PF02385; OMP; 1.
CC KW Outer membrane; Signal; Multigene family.
CC FT SIGNAL 1 25 POTENTIAL.
CC FT CHAIN 26 928 PROBABLE OUTER MEMBRANE PROTEIN PMP10.
CC SQ SEQUENCE 928 AA; 97229 MW; 0590D5206A1DD0E1 CRC64;

```

```

Query Match 2.9%; Score 115; DB 1; Length 928;
Best local similarity 24.2%; Pred. No. 0.72;
Matches 107; Conservative 39; Mismatches 145; Indels 152; Gaps 23;

```

```

QY 338 VLGNOPDIYWDLSAVHGLGLVHPKVFNFYNELHAYLASCGVGVKVDVONITETLGAAGH 397
Db 271 ISGQSVTFESGNOAVANGAIVAKKL-----TLASGGGGISFS--NNIVOGTTAGN 320
QY 398 GGRVSLTRSYHHALEASISNFTDNCIACMCNHTDGLYSAKOTAIVRASDDFPRDPAS 457
Db 321 GGAISILAAEGCSLSAE-AGDITFNG-----NAIVATY--PQTRK 358
QY 458 HTTHISSAVNSLFLGEMQPDWDMFRSLHPADYHAARAIGGCPIYVSDKRGNNRPL 517
Db 359 NSIDIGSTAKIT-----NLRAISGHSIF----- 381
QY 518 LKKVLPLPDSVLAQLPGRPTDLSFVDP-----ARDRTSLIKWMLNKC-----SG 564
Db 382 -----PYDPTTANTADSDTL--NLNKAADAGNSTDYSG 413
QY 565 VVGWFCOGAGWCKIEKTRIHDT-----PGLTAS--VCASDVLD-----ITOVAGAEML 614
Db 414 SI-VES--GKLSDEDAKADVADNLSTLKPVTLLAGMLVLRGVTLDTGKFTOTAGSSVI 470
QY 615 GDTIVVVRSGEVYRLPKGSIVPT-----LKVLEPELHFHCYIQELAPISFAAG-- 666
Db 471 MDAQTLKASTEEVTL-TGHSIPVDSIGEGKKV-----IASASAKNALVLSGPII 520
QY 667 LLDGFNTGAVEQVEIHNRATKTIALSVGRGRFGVSSQRPDKCVGGAETDFNYDSE 726
Db 521 LLD--NQGNAVENHDLKTDQDFVQLSALGTAT-----TIDVAVPPTV-ATPHNGVGGT 573
QY 727 TGLT-----TESIPVSPDEMYRWS 745
Db 574 WGMTWVDDTASTPKTKTATLAWT 596

```



```

QY 435 LSKMGTATVRSADDDPYRD-----PASHIHTSSVANSPLGEF 475
Db 428 VETGK--SYLKANGIDIPDEGLALAGLGEKGVSPDLQAGFHTFAANGTYTEPPISSTI 485
QY 476 MGPDP-----WDMHSIHPADYHAARAIGGCPIVSDKPNHNF 515
Db 486 IDDEGETIADHKKEGGRVFSKOTSNMTRMLQYVKKGTAT-----SOTYHG 532
QY 516 DLKLKVLPGDSYLRQQLPGRPTRDLSLEVPDARDRTSLIKIWMLNKSGVGVENCQAG 575
Db 533 DLNGK-----TGSTSYGVSG-ATKQAMPAGYTPKIT-----GAV 566
QY 576 WCKIEKTRTRHDPSTPTLASFVSCADVDLITOVAGAEMLGPTIYVARSGEVIRLPGV- 634
Db 567 WMGYDKTDQNHIIYAKGSSPYRLFDIT--LQA-----GET-----GHVTFKKVKK 611
QY 635 --SIPVTLKYLEFELPHFCPIQELAPISISPAIGL-----DMFNTGAVEQVEIHNRA 686
Db 612 ELSESPLEK-----PVKTLTADYTFKAAGLFTLELKWDAQEDDRAVYRIYV-NKD 660
QY 687 ATKTIALSVRGREGV 703
Db 661 GEETLDSVVGKGYEI 677

RESULT 6
PMPE.CH1MU STANDARD: PRT: 976 AA.
AC 09p147;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE -Probable outer membrane protein pmpe precursor (Polymorphic membrane protein E).
GN PMPE OR TC0261.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83360;
[1]
SEQUENCE FROM N.A.
RP STRAIN-MOPn / Ni95;
RC MEDLINE=20150255; PubMed=10684935,
RX Read T.D., Brunham R.C., Shen C., Gail S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gysin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1. SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -1. SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on lists
CC used by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AE002293; AAF39130.1; -
CC TIGR: TC0261; -
CC InterPro: JPRO03357, OMP.
CC Pfam: PF02385; OMP; 1.
CC Outer membrane; Signal; Multigene family; Complete proteome.
CC SIGNAL 18 POTENTIAL.
CC CHAIN 19 976 PROBABLE OUTER MEMBRANE PROTEIN PMPE.
CC SEQUENCE 976 AA; 105964 MW; 0BEA56158E3CEA9 CR64;

```

	Best Local Similarity	18.5%: Pred. No. 3-9:	Matches 116:	Conservative 73:	Mismatches 183:	Indels 254:	Gaps 29:
QY	104	SETDGENSPIITYVLLPLLEGGFRAVLQGNDRKNEIEICLEGSDNAVEFDGL-----HMV	158				
Db	139	TDYDPEP-----EENKARE-GGAIHAGDYIISNNQNLVGFIKNFA	177				
QY	159	YMAHGTVPEFVINOAKVAEKKMQFFLHHEKKRLPSCLDMGCKTWIAFTDVTAESEVE	218				
Db	178	YVGG-----AISA-----NPFAYKENT-----	195				
QY	219	GKSLSGGTPRPFLIIDGWOQIEKKAKDAECYVGAQAFATRLGIEKNTFOKIQ	278				
Db	196	-----SSFLQNNSCIQYTKTGKG-----GAIYVTSOSFENKKDLFLIQ	236				
QY	279	NNEQMSGKLVLHGAKOHNNKKNVYVMAALAGYGGVKAPEATGKEHYDTALAPVQSPCY	338				
Db	237	N-----SCCAGG-----AIFSPQCS--L	252				
QY	339	LGNDPIVMDSLAVHGLGVHPKRYEFENYELHAYLASCVDYGVQYONIEITGAGHG	398				
Db	253	IGNQGDIVFES--NHGF-----KNDNANNE-----SGDS	280				
QY	399	GRVSLRFSYHHALEASIASN-----FTDNGCIACCHNTDGLYSAKQATVIRASDDEYPR	453				
Db	281	GAIVLT-----TRLDITNNGSQIFESFD-----ISRNFGAIHAPCLHLVNGNPTFTN	329				
QY	454	DPASHT---HHISSVAVNSLFLGEFNPQPMQMFHSLHRAALYHA-----AARATIGCP	503				
Db	330	NIANHTGCAIYITGTETSKI-----SADHMAIIPEDNNISANATATAGG	371				
QY	504	IYVSDKEGNHNFDLKKLVLPDGSVLRAQLPGRPRDSLFEVDPARDTSLKIMWLNKCS	563				
Db	372	SSSTNPENPHNN-----ATIMDSAGIGELGAKGQNLIFPDQVNTAGVT--DEKKA	424				
QY	564	GVCVFEGVQCQAGCKITE-KKTRIHDSPEPTLAS---VCASD-----VDLTQYAGAEWL	614				
Db	425	SQTCGVFSQATVYLSADISQANLQTKTPATVLTLSGCLLCIEDRAQLVTNNPTGGGIVAL	484				
QY	615	GDITV---VAYRSGEVRILPK-----GVSIPTVLK-VLEFELFHRCPI---	653				
Db	485	GNQAVLSSYQHSHTDANTQRPPTTTTDDASVLTNHTGMLPRLINDGAEMLPWEPISTT	544				
QY	654	-----QELAPISIF--AAIGLDD	669				
Db	545	QGNTTYTSDTAASFSLNGATISLID	570				
RESULT	7						
PRB_RAT		STANDARD:	PRT:	555	AA.		
ID	Q00438:	Q63568:					
AC	01-APR-1993 (Rel. 25, Created)						
DT	01-APR-1993 (Rel. 25, Last sequence update)						
DT	01-APR-1993 (Rel. 25, Last sequence update)						
DE	16-OCT-2001 (Rel. 40, last annotation update)						
DE	poly(pyrimidine tract-binding protein (PUB) (Heterogeneous nuclear ribonucleoprotein I) (hnRNP I) (Pyrimidine-binding protein) (PYBP).						
GN	PUB OR PYBP OR IFRPI.						
OS	Rattus norvegicus (Rat).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.						
OK	NCBI_Taxid=10116;						
RN	[1]						
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 185-205; 348-365 AND 517-549.						
RC	TISSUE=Hepatoma;						
EX	MEDLINE=92020211; PubMed=1681508;						
RA	Binuel F., Alzari P.M., Ferrara P., Zakin M.M.;						
RT	"Cloning and sequencing of PYBP, a pyrimidine-rich specific single strand DNA-binding protein.";						
RL	Nucleic Acids Res. 19:5237-5245(1991).						
RA	Sequencia P.,						

RL Submitted (Aug-1993) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: PLAYS A ROLE IN PRE-MRNA SPLICING. BINDS TO THE
 CC POLYRIMIDINE TRACT OF INTRONS. MAY PROMOTE THE BINDING OF U2
 CC SNRNP TO PRE-MRNA.
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: PYBP1 AND PYBP2
 CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC - SIMILARITY: CONTAINS 4 RNA RECOGNITION MOTIFS (RRM).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X60789; CAA43202.1; -
 CC EMBL: X60790; CAA43203.1; -
 CC EMBL: X74565; CAA52653.1; -
 CC InterPro: IPR00504; RRM.
 CC Pfam: PF00076; RRM; 4.
 CC SMART: SM00360; RRM; 4.
 CC PROSITE: PS50102; RRM; 4.
 CC PROSITE: PS00030; RRM_RNP_1; FALSE NEG.
 CC Nuclear protein; RNA-binding; mRNA splicing; Repeat;
 CC Alternative splicing.
 CC DOMAIN 58 142 RNA-BINDING (RRM) 1.
 CC DOMAIN 183 259 RNA-BINDING (RRM) 2.
 CC DOMAIN 361 435 RNA-BINDING (RRM) 3.
 CC DOMAIN 478 553 RNA-BINDING (RRM) 4.
 CC DOMAIN 340 347 POLY-ALA.
 CC VAAPSLIC 297 321 MISSING (IN ISOFORM PYBP1).
 CC CONFLICT 57 57 S -> T (IN REF. 2).
 CC CONFLICT 164 164 A -> R (IN REF. 2).
 CC CONFLICT 311 320 VPSHCHPSR -> GPEPTAIPAQ (IN REF. 2).
 CC CONFLICT 542 549 ENHILRVS -> LGDNHHR (IN REF. 1; AA
 CC SEQUENCE).
 CC FT SEQUENCE 555 AA; 59353 MW; E73F22B5467117F CRC64;
 CC SQ
 Query Match 2.6%; Score 105.5; DB 1; Length 555;
 Best Local Similarity 18.8%; Pred. No. 1.9;
 Matches 110; Conservative 77; Mismatches 216; Indels 181; Gaps 27;
 OY 10 NDGKLVHGKTLTGVPDNYVLTGSGGLYTGAFVGTASHKSLHFPVGLGLEGRFM 69
 DB 41 NDSK-KFKGDNSTGVPSRYI-----HVRKL--PSDVTGG----- 72
 OY 70 CCFRKLMMWQRMGTCGNDVPLETQFMLESKE-----SETDGENSPI-IYVLLPL 122
 DB 73 -----EVLISGLPCKRYTNLMKCKNKAFTEMSEEAANMVMYTSVAVYL 120
 OY 123 EGSFRAVLQGNKNEIEICLESQNAVETDGLAHVYHAGTNPPEVINQAVKAVEKHMQ 182
 DB 121 RQPIIYIGPSNMK---ELKTDSSPNQARAQAALQAV-----NSYQSGNLALAA----- 165
 OY 183 TELHREKRLPSCIDMFQMGCTDADFYDTVATGVEGGLKSLSGSGTTPRRLIIDGQMOI 242
 DB 166 -----SAAVADAG--MMAQOSPLRLT-----V 187
 OY 243 ENKAKDAT-ECLVQEGAPAT--RLTGKENTKFOKLIQNNQOMGLKHLVGAQHNNV 299
 DB 188 ENLFTVYITDVLHOLFQSKGTGVLKITTFFKNNQFQALQYADPVGA-QH-----ATLSLDG 242
 OY 300 KNYVYVHALA-----GYWGKVPAAATGMEHDTALATVYQSPG 337
 DB 243 QNYI--NACCTLRIDFSKLTSLNVKYNNDKSRDTPRPDPSGDSQPSLDQTAFAFGAPG 300
 OY 338 VLGNDIDYMDSLAVHGLGLVHPKKVFNYNELHAVILASCGDGV-----KYDVQNI 389
 DB 301 IMSASP--YAGAVPSH--LCHPSRAGLSVPNVHGALADLALPSAAAAAAGRAITFGL 355

OY 390 IETLGAGHGRVSLTRSYH--HALEASIASNFTDNGCIACMCNHTDGLYSAKQATVRA 447
 DB 356 -----AGAGNSULVSLNINPERVPPQSILFEGYGDV-----QRYKILFNKENALVEMA 406
 OY 448 DDFYRDPASP-----TIHSSVAVNSLFLGEPQPMWDFHSLHAPADVHAARA 498
 DB 407 DGSQAOAMSHLNGHKLHGKSVRTLSKHOSVOLPREGQEOGL-----TRDY----- 454
 OY 499 ICGCPHYSDKPGNNHNPDLKLVLDGQSVLRAQDGPRTDGL 542
 DB 455 -GSSPLHRRKPKGSKNF--QNIFFPSATIHLSNIPSVSDDL 494
 RESULT 8
 ID P1CP_PSESR STANDARD; PRT; 587 AA.
 AC P42790;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Pseudomonaspepsin precursor (EC 3.4.23.37) (Pepstatin-insensitive
 DE carboxyl proteinase).
 GN PCP.
 OS Pseudomonas sp. (strain 101) (Achromobacter parvulus T1).
 OC Bacteria; Proteobacteria.
 OX NCBI_TaxID=33067;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 216-224.
 RX MEDLINE=95014496; PubMed=7929375.
 RA Oda K., Takahashi T., Tokuda Y., Shibano Y., Takahashi S.;
 RT "Cloning, nucleotide sequence, and expression of an isovaleryl
 RT pepstatin-insensitive carboxyl proteinase gene from Pseudomonas sp.
 RT 101.";
 RL J. Biol. Chem. 269:26518-26524(1994).
 RN [2]
 RP SEQUENCE OF 216-585.
 RX MEDLINE=96157710; PubMed=8576087;
 RA Hayashi K., Izu H., Oda K., Fukushima K.-I., Matsuo M., Takano R.,
 RA Hara S.;
 RT "The primary structure of pepstatin-insensitive carboxyl proteinase
 RT produced by Pseudomonas sp. No. 101.";
 RL J. Biochem. 118:738-744(1995).
 RN [3]
 RP MUTAGENESIS OF ASPARTIC ACID AND GLUTAMIC ACID RESIDUES, AND ACTIVE
 RP SITES.
 RX MEDLINE=99419069; PubMed=10488127;
 RA Oyama H., Abe S.-I., Ushiyama S., Takahashi S., Oda K.;
 RT "Identification of catalytic residues of pepstatin-insensitive
 RT carboxyl proteinases from prokaryotes by site-directed mutagenesis.";
 RL J. Biol. Chem. 274:27815-27822(1999).
 CC - CATALYTIC ACTIVITY: Hydrolysis of the B chain of insulin at 13-
 CC Glu-1-Ala-14, 18-Leu-1-Tyr-19 and 25-Phe-1-Tyr-26 and angiotensin
 CC I at 4-Tyr-1-Ile-5. A good synthetic substrate is Lys-Pro-Ile-Glu-
 CC Phe-1-Phe(NO2)-Arg-Leu.
 CC - SUBCELLULAR LOCATION: Periplasmic.
 CC - PTM: AUTOCATALYTICALLY PROCESSED.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S53.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D37970; BAA07188.1; -
 DR MEKOPS; S53.001; -
 KW Hydrolyase; Protease; Serine protease; Zymogen; Periplasmic; Signal.
 FT SIGNAL 1 215 POTENTIAL.
 FT PROPEP 2 215 REMOVED IN NATURE FORM.
 FT CHAIN 216 585 PSEUDOMONASPEPSIN.
 FT PROPEP 586 587 REMOVED IN MATURE FORM.

FT ACT_SITE 385 385 PROBABLE.
 FT ACT_SITE 502 502 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 543 543 PROBABLE.
 FT DISULFID 352 391
 SO SEQUENCE 587 AA; 61072 MM; E193DB82C225829A CRC64;

Query Match 2.6%; Score 103.5; DB 1; Length 587;
 Best Local Similarity 21.8%; Pred. No. 2.9; Mismatches 204; Indels 159; Gaps 26;
 Matches 114; Conservative 47; Mismatches 204; Indels 159; Gaps 26;

QY 243 ENKAKATECLVOEG-AQFATRLTGKENTKQKLLQNNQMSGLKHLVGAQH-----296
 DB 82 EAOLKOLAQAVNPGNAQCG-----KFLKRRQFLSOFAPTAQOAVAHLRKNG 131
 QY 297 -----HMKVNYVHVALGHWGVKPA-ATGMEHYD-----ALAVPYSPGVLGNDPDI 345
 DB 132 FVAIHVYVPRNLISA-DGSAQAVKAFNTPVRYQLNGKAGYANAPADVPDLDG---EI 187
 QY 346 VMDSLAVHGLVHPKRVENFYNELHAYLASGVDSVDKVDYONIIETLGGHGRVSLFR 405
 DB 188 VGSVLGLQNVTRAHP-----LKGERSAKTLAAG-----TA 220
 QY 406 SYHHALEASTASFTDNGCTACKCHNTDGLYSAKQTAIVRASDDFYPDPASHTIHSSV 465
 DB 221 KGNHPTFEPT-----IYDASSAPTA-----AMTVGIITR 250
 QY 466 AVYSLFGEFMDPMDHSLHPADYHAAARAIGGC-PIYVSDKRGNNIFLLKVIPL 524
 DB 251 GGVSQLQLDLOQ-----FTSANGLASVNTQTQTGSSNGSDYDDQGGQEWLDSQSTIV- 304
 QY 525 DGSVLAQALPGRTDLSLFV---DPAQRTSLKIMLNKSCGVGVGVCQAGACMKIEK 581
 DB 305 -SSA-----GGAVOQLLFYMAQDQASGNTGLQAFNQAVSDNNVAKVINV-SLGCCEADA 356
 QY 582 KRIHDSRGTLTASVCASDVLLITQVGAEM-----LGDITIV-----AYRSEVITRL 630
 DB 357 -----NAGDTLQAE-----DRIFATAAQQOTFSVSSGDEGVYECUNRGYDGSSTV-- 402
 QY 631 PKGSIPTVTLKLEFELHFCPILOELAPISFAIIGLDMFT--GGAVEQVETIHNRAIK 689
 DB 403 --SVSNP-----ASSPNVIAVGGTTLVYTSAGAAVSNETVWNEGIDS 441
 QY 690 TIALSVRGKRGFGVYSSQRLKCVVGAEE---TDFVDSSETG 728
 DB 442 NGKLWATG-GGYVYESKPSQSVSGTPGRRLPDISPDAAG 484

RESULT 9
 RPL_EBOSM STANDARD: PRT: 2210 AA.
 ID RPL_EBOSM 06802;
 AC 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE RNA-directed RNA polymerase (EC 2.7.7.48) (Large structural protein) (L protein).
 GN L.
 OS Ebola virus (strain Sudan Maleo-79) (Ebo).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
 CC Filovirus.
 NCBI_TaxID=128949;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sanchez A., Trappier S., Nichol S.T.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACTIVE POLYMERASE. IT
 MAY FUNCTION IN RNA SYNTHESIS, CAPING, AS WELL AS METHYLATION OF
 CAPS, AND POLY(A) SYNTHESIS. ALSO INVOLVED IN TRANSCRIPTIONAL RNA
 EDITING.
 CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 (RNA)(N).
 CC -I- SIMILARITY: PRIMARILY WITH THE N-TERMINAL HALF OF THE L PROTEINS

CC OF RHABDOVIRUSES AND PARAMYXOVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U23458; AAA79970.1;
 DR InterPro: IPR001016; Paramyx_RNA_pol.
 DR Pfam: PF00946; Paramyx_RNA_pol.
 KW transferase, RNA-directed RNA polymerase.
 SO SEQUENCE 2210 AA; 251255 MW; F50E6B624951AE44 CRC64;

Query Match 2.6%; Score 103.5; DB 1; Length 2210;
 Best Local Similarity 22.2%; Pred. No. 20; Mismatches 195; Indels 123; Gaps 25;
 Matches 107; Conservative 56; Mismatches 195; Indels 123; Gaps 25;

QY 315 VKPATGMEH--YDTALAVPYQSPGVLGNDPDIYMDSLAVHGLVHP---KKVFNFYN 368
 DB 483 IKDRATAVEQTCMDA-----VEEPNVLGSPRYFNRKRVPEQFLQEDDSIESVLYQAQ 537
 QY 369 ELHAYLASGVDSVDKVDYONIIETLGGHGRVSLTSSYHHALEA---SIASFTDNGC 424
 DB 538 ELRYLLPQNNRFSFLKEREK--NVGRTEGKLPYLITRVNVTICBALLADGLAKAFPSNM 595
 QY 425 IACMCHNTDGLYSAKQTAIVRASDDFYPDPASH-TIHSSVA-----YNSLFGEMQ 478
 DB 596 VYTEREKESL--LHQASMHHTSDP-----GEATVYRGSSFYDLEKYNLAERFETAP 648
 QY 479 -----DMDHSHLHPADYHAAARAIGCCPIYVSD--KPSNNHFDLLK 521
 DB 649 FIKYCNCYGVANVEDW--MHFLIPQ-----CYMHVSDYVNP--PHNVTLENRE 693
 QY 522 VLPRG-SYLRQALPG-KPTDLSLFVDPARDRTSLKLT-----WNLNKSCGVVGV 569
 DB 694 YPPGSAVYRGHGLGLOQKMTSISCAQISLVEIKTFKLSAVAGNDQCITVLSVF 753
 QY 570 NCQAGCMCKIEKRIHDSRGTLTASVCASDVLLITQVGAEMLGD-TIIVAYRSGEVI 628
 DB 754 PLESSP-----NEQRCADENARVAASLAK-----VISACGIPLKPDETVH---SGFIT 801
 QY 629 RLKP-----GVSIPTLVK-----LEFELP-----HFCPIQEL 656
 DB 802 FGPKQYINGIQLPQSLKTAARMAPLSDAIFPDLOGTLASIGTAFERSISFTRHILPSRVA 861
 QY 657 APSISFAIIGLDM-----FMTGCAVEQVEIHNRAATITLAVRGKRGFRVYSSQRLK 712
 DB 862 AAFITFSVRILHGHGIFKHSGLQGLALNKPLDFETLALAVPVLGSLFLNPERK 921
 QY 713 V 713
 DB 922 L 922

RESULT 10
 POLG_POL3L STANDARD: PRT: 2206 AA.
 ID POLG_POL3L
 AC P03302; Q84783; Q84784; Q84785; Q84786; Q84787; Q84788; Q84789;
 AC Q84790; Q98592; Q98593; Q98594;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
 DE P2A TO P2C; P3A; Genome-linked protein VEG; Picornain 3C
 DE (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
 DE (EC 2.7.7.48)].
 OS Poliovirus type 3 (strains P3/Leon/37 and P3/Leon 12A11B).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 CC Enterovirus.

OX NCBI_TaxID=12088;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P3/LEON/37;
 RX MEDLINE=84170338; PubMed=6324200;
 RA Stranway G., Hughes P.J., Mountford R.C., Reeve P., Minor P.D.,
 RA Schild G.C., Almond J.W.;
 RT "Comparison of the complete nucleotide sequences of the genomes of
 RT the neurovirulent poliovirus P3/Leon/37 and its attenuated Sabin
 RT vaccine derivative P3/Leon 12a1b.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:1539-1543(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P3/LEON 12A11B;
 RX MEDLINE=83299239; PubMed=6310508;
 RA Stranway G., Cann A.J., Hauptmann R., Hughes P.J., Clarke L.D.,
 RA Mountford R.C., Minor P.D., Schild G.C., Almond J.W.;
 RT "The nucleotide sequence of poliovirus type 3 Leon 12 alb: comparison
 RT with poliovirus type 1.";
 RL Nucleic Acids Res. 11:5629-5643(1983).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-878.
 RX MEDLINE=95120467; PubMed=7820548;
 RA Grant R.A., Hirmath C.N., Filman D.J., Syed R., Andries K.,
 RA Hogle J.M.;
 RT "Structures of poliovirus complexes with anti-viral drugs:
 RT implications for viral stability and drug design.";
 RL Curr. Biol. 4:784-797(1994).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1-878.
 RA Syed R., Filman D.J., Hogle J.M.;
 RT Submitted (MAR-1995) to the PDB data bank.
 RL
 CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 CC Q/Q SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PPM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: THE SEQUENCE OF STRAIN SABIN VACCINE P3/LEON/37 IS
 CC SHOWN.
 CC -1- MISCELLANEOUS: THE STRAIN SABIN VACCINE P3/LEON/37 IS THE
 CC PROGENITOR OF THE STRAIN SABIN VACCINE P3/LEON 12A11B.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL: K01392; AAA46914.1; -
 DR EMBL: X00925; CA25444.1; -
 DR PIR: A03900; GNNY4P.
 DR PDB: 1PIV: 20-JUL-95.
 DR PDB: 1PVC: 15-SEP-95.
 DR PDB: 1YBA: 11-JUL-96.
 DR PDB: 1YBB: 11-JUL-96.
 DR PDB: 1YBC: 11-JUL-96.
 DR PDB: 1YBE: 11-JUL-96.
 DR MEROPS: C03.020; -
 DR InterPro: IPR001199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rny.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.

DR Pfam: PF00073; Rny; 3.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KW Polypeptide; Coat protein; Core protein; Transferase;
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
 KW 3D-structure.
 FT CHAIN 2 69 COAT PROTEIN VP4.
 FT CHAIN 70 340 COAT PROTEIN VP2.
 FT CHAIN 341 578 COAT PROTEIN VP3.
 FT CHAIN 579 878 COAT PROTEIN VP1.
 FT CHAIN 879 1027 CORE PROTEIN P2A.
 FT CHAIN 1028 1124 CORE PROTEIN P2B.
 FT CHAIN 1125 1453 CORE PROTEIN P2C.
 FT CHAIN 1454 1540 CORE PROTEIN P3A.
 FT CHAIN 1541 1562 GENOME-LINKED PROTEIN VP6.
 FT CHAIN 1563 1745 PICORNAIN 3C.
 FT CHAIN 1746 2206 RNA-DIRECTED RNA POLYMERASE P3D.
 FT LIPID 2 MYRISTATE.
 FT ACT_SITE 1709 1709 PROTEASE (POTENTIAL).
 FT ACT_SITE 1723 1723 PROTEASE (POTENTIAL).
 FT VARIANT 431 431 S -> F (IN P3/LEON 12A11B).
 FT VARIANT 864 864 K -> R (IN P3/LEON 12A11B).
 FT VARIANT 908 908 T -> A (IN P3/LEON 12A11B).
 SO SEQUENCE 2206 AA; 246163 MW; 4766B15C861F66D3 CRC64;

 Query Match 2.6%; Score 102; DB 1; Length 2206;
 Best Local Similarity 21.0%; Pred. No. 27;
 Matches 122; Conservative 67; Mismatches 207; Indels 184; Gaps 33;

 QY 10 NDGKLVH-GKTLITGVPPNVVLTG-----SGR-----GLY 40
 DB 1660 NDGVLLVNSKRYNMVPGVAVTEQGYLNLGNGQTARILMYNPPRAGCGGVITCTGV 1719
 QY 41 TGAFTGATASHSKSLHVPFGVLEGLREKCCERFKIMMTORMGTCGRVPLETQPMLE 100
 DB 1720 IGHWVGNGSH-----GFAAL-----KRYSTQSG-----ETQMR-P 1753
 QY 101 SKSEFDGENSPITLYVLPLE-GQFRAVLQNDKNEIEICSGDNAVETDGLHWY 159
 DB 1754 SKRA-----GYPLINAPTKLPSAFHYFEVKEPAV-----LTKNDPRLKIDFEALIS 1805
 QY 160 MHAGTNPFEYINQAVAKVEKMQTFLHREKKRLPSCLDFGWCWTDAFY-TDVTAEGVEE 218
 DB 1806 KYVGKNTKEVDEYKMEKVDHYAGQLMSLDISTBQMLE-----DAMYGTD----- 1850
 QY 219 GLKSL--SGCGTPRFLIIDGWOQLEK---ADATBCLVQEGAOAFATRLTGIKENT 271
 DB 1851 GLRALDLSTAGYR--YVAMGKKRRDLINKOTRDTREKORLLDAYGINLP-VTVYKDEL 1907
 QY 272 KPKKCLONNQMGSGLKLVGAKOHNVVYVWHAAGVGVKPAATGMEHYDTALAY 331
 DB 1908 RSKTKVE-----QKSKLIIASSLSNV-----AMRAFGLYNAF-----H----- 1944
 QY 332 PVQSPGV-----LGNPDYVMSLAVHGLGLVHPKKVFN-YWELHAYVASCVDGVKVD 385
 DB 1945 --RNPGVVTSAGVCDPDLFWSKIPV-----LMEKILFADYGYDASLSPAMFEALKNV 1997
 QY 386 VONTIFELGAGHGRVS-----LRSYHHALEAIASTFDTNGCIACMCHTDLVSAKQT 441
 DB 1998 LEKI-----GFGDRVDYIDLHNS-HILYKNKI--YCVKGGMPSCSGSTISFNSMINN 2047
 QY 442 AIYVA-----SDPE-----YPRD-----PASHTH 461
 DB 2048 LITLTLLKTYKGIIDDLKMAVGDVYASVYHVDASLSAGSKDYGLTMPRPAKSAFT 2107
 QY 462 ISSVAY-NSLFLGEMQPDWD---MFHSLHPADYNAAR 497
 DB 2108 FEVTVWENVTFLKRFRADEKYPFLIHPVMPKKEIHSIR 2147

CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- CAUTION: SEE ALSO VERSION 1 OF THIS PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: V01149; CAA24461.1; ALT_SEQ.
DR EMBL: V01149; CAA24462.1; ALT_SEQ.
DR EMBL: V01149; CAA24463.1; ALT_SEQ.
DR EMBL: V01149; CAA24464.1; ALT_SEQ.
DR PIR: A03898; GNNY2P.
DR PDB: 2PLV; 15-JUL-93.
DR PDB: 1FPT; 31-MAR-95.
DR PDB: 1POV; 07-DEC-95.
DR PDB: 1VBD; 11-JUL-96.
DR PDB: 1AL2; 19-NOV-97.
DR PDB: 1AR6; 03-DEC-97.
DR PDB: 1AR7; 03-DEC-97.
DR PDB: 1AR8; 03-DEC-97.
DR PDB: 1AR9; 03-DEC-97.
DR PDB: 1ASJ; 03-DEC-97.
DR PDB: 1PO1; 03-DEC-97.
DR PDB: 1PO2; 03-DEC-97.
DR MEROPS: C03.020. -.
DR -
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR003138; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rhv_3.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF00073; rhv_3.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
KW Polypeptide; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
KW 3D-structure.
FT INIT MET 0 0
FT CHAIN 1 68 COAT PROTEIN VP4.
FT CHAIN 69 340 COAT PROTEIN VP2.
FT CHAIN 341 578 COAT PROTEIN VP3.
FT CHAIN 579 880 COAT PROTEIN VP1.
FT CHAIN 881 1029 CORE PROTEIN P2A.
FT CHAIN 1030 1126 CORE PROTEIN P2B.
FT CHAIN 1127 1455 CORE PROTEIN P2C.
FT CHAIN 1456 1542 CORE PROTEIN P3A.
FT CHAIN 1543 1564 GENOME-LINKED PROTEIN VP6.
FT CHAIN 1565 1746 PTCORNAIN 3C.
FT CHAIN 1747 2208 RNA-DIRECTED RNA POLYMERASE P3D.
FT LIPID 1 1 MYRISTATE.
FT ACT_SITE 1711 1711 MYRISTATE.
FT ACT_SITE 1725 1725 PROTEASE (POTENTIAL).
FT STRAND 3 6
FT STRAND 25 28
FT STRAND 35 37
FT HELIX 45 45
FT STRAND 49 49
FT TURN 49 49

FT STRAND 50 53
FT HELIX 56 56
FT TURN 62 63
FT TURN 74 75
FT TURN 79 80
FT STRAND 82 86
FT TURN 87 88
FT STRAND 89 93
FT STRAND 96 101
FT HELIX 102 104
FT TURN 112 114
FT STRAND 122 122
FT HELIX 125 127
FT TURN 128 128
FT STRAND 132 133
FT STRAND 137 140
FT TURN 141 142
FT STRAND 146 150
FT TURN 151 151
FT HELIX 152 154
FT TURN 155 156
FT HELIX 158 166
FT STRAND 167 179
FT TURN 184 185
FT STRAND 186 196
FT TURN 197 197
FT STRAND 202 202
FT HELIX 213 216
FT STRAND 219 221
FT STRAND 223 224
FT STRAND 226 226
FT TURN 235 236
FT STRAND 242 242
FT STRAND 245 245
FT HELIX 246 248
FT TURN 249 252
FT HELIX 255 260
FT STRAND 263 267
FT TURN 268 270
FT STRAND 273 278
FT TURN 287 287
FT STRAND 289 291
FT STRAND 292 292
FT STRAND 295 307
FT STRAND 314 331
FT STRAND 337 337
FT STRAND 348 351
FT TURN 355 355
FT STRAND 363 363
FT TURN 366 367
FT STRAND 379 380
FT STRAND 382 382
FT HELIX 383 387
FT TURN 388 388
FT STRAND 391 392
FT STRAND 397 397
FT TURN 399 403
FT HELIX 405 408
FT STRAND 410 413
FT STRAND 423 426
FT TURN 429 431
FT TURN 433 437
FT HELIX 439 444
FT TURN 445 446
FT STRAND 447 451
FT STRAND 454 460
FT TURN 464 465
FT STRAND 467 467
FT STRAND 469 475
FT HELIX 485 489
FT TURN 490 490
FT STRAND 492 497
FT STRAND 503 508

FT STRAND 517 518
 FT HELIX 523 525
 FT STRAND 529 534
 FT STRAND 539 539
 FT TURN 542 543
 FT STRAND 547 556
 FT TURN 558 559
 FT STRAND 561 565
 FT STRAND 600 600
 FT TURN 601 602
 FT STRAND 604 604
 FT STRAND 608 608
 FT TURN 620 621
 FT STRAND 622 623
 FT HELIX 625 627
 FT HELIX 635 637
 FT TURN 638 638

Query Match 2.5%; Score 101; DB 1; Length 2208;

Best Local Similarity 19.8%; Pred. No. 32; Mismatches 214; Indels 176; Gaps 29;

Matches 114; Conservative 72; Mismatches 214; Indels 176; Gaps 29;

QY 10 MDGKLVYH-----GKTLTGVDNVYLPFG--SGRGV 40
 DB 1662 MDGVLIVNTSKYPMNVYVGAETEGCYLNLGGRTAFRLMNPFRAGCGGVITCGKV 1721

QY 41 TGAFAVATASHSKSLHVFPMGLBGLRMCCFRKLMMTQRMGTGGRDVELEQPMLE 100
 DB 1722 IGHNVGNGSH-----GFAAL-----KRSTQSOG-----EIQWR-P 1755

QY 101 SKSESTDGNSPIIYTVLLPLE-GQFRAVLQGNKNEIEICSGDNAVEFDGLHMV 159
 DB 1756 SKS-----VGPLINAPSKTKLEPSAFHYVEGVKEPAV---LKNDRKLTDDEEALFS 1807

QY 160 MHAGTPEFVINQAKVAKVKNQOTLHREKKRLPSCLDLFGKCTWDAFY-TDVTAGVEE 218
 DB 1808 KYVGNKITEVDKMEADYHAGLSLDINTDQCLD-----DAVYGTD----- 1852

QY 219 GKSL-----SQGSTPREFLIIDGMOIENK-----AKDATECLVQGAQFATRLGKENT 271
 DB 1853 GLEADLSTAGYR--YVAMGKKRRDILNKQTRDKENQKLDLDTYGINLP-VTVVDEL 1909

QY 272 KPOKLIQNEQMSGLKHLVHGAKOHNNKVVYVHALAGYGVKPAATGMEHTDALAY 331
 DB 1910 RSKTYE-----QGSRLLEASLNDV-----AMEMAGNITYAAV---H----- 1946

QY 332 PVQSPGVL-----GNQPDVMDSLAVHGLGVHPKKVNF-YNELHAYLASCVDGVKVD 385
 DB 1947 --KNQVITIGSAVGCDDPFLFMSKIPV-----LMEKILFAFDYTGVDASLSAFMEALKV 1999

QY 386 VONIETLGAGHGRVSLTRSYHNALEASINFTDNCIACMCHNTDGLYSARQTAIVR 445
 DB 2000 LEKI-----GFDVYDYIDYLNHSHNLKKNKYCVKGMPSGSGSIFSMINNLIR 2053

QY 446 A-----SPDF--YPRD-----PASHTIHLSV 465
 DB 2054 TLLTKYKCIDDLKMLAYGDDVYASRPHVEDASLSAQSGKDYGLTMTPADKSANPEVY 2113

QY 466 AY-NSLIFGEFQMPDMD--MFHSLRPADYHAAR 497
 DB 2114 TWENVTLKRFRADEKYPFLIHVPMKKEIHESIR 2149

RESULT 14
 DYHC_ANTCR STANDARD; PRT; 4466 AA.

AC P39057;
 DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Dyein beta chain, ciliary.
 OS Anthocidaris crassispina (Sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

CC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Echinometridae;
 OC Anthocidaris.
 NCBI_TaxID=7629;
 OX [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=91326104; PubMed=1830928;
 RA Ogawa K.;
 RT "Four ATP-binding sites in the midregion of the beta heavy chain of
 dynein.";
 RL Nature 352:643-645(1991).
 CC -I- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
 FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
 CC DYEIN HAS ATPASE ACTIVITY.
 CC -I- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA),
 CC THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
 CC -I- SIMILARITY: BELONGS TO THE DYEIN HEAVY CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL, D01021; BA00827.1; -.
 DR PIR, S17231; S17231.
 DR InterPro; IPR004273; Dynein_heavy.
 DR Pfam; PF03028; Dynein_heavy; 1.
 DR Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
 KW Coiled coll.

FT DOMAIN 482 502 COILED COLL (POTENTIAL).
 FT DOMAIN 627 643 COILED COLL (POTENTIAL).
 FT DOMAIN 734 805 COILED COLL (POTENTIAL).
 FT DOMAIN 1036 1056 COILED COLL (POTENTIAL).
 FT DOMAIN 1306 1337 COILED COLL (POTENTIAL).
 FT DOMAIN 1443 1468 COILED COLL (POTENTIAL).
 FT DOMAIN 1950 1978 COILED COLL (POTENTIAL).
 FT DOMAIN 3033 3134 MICROTUBULE-BINDING (POTENTIAL).
 FT DOMAIN 3263 3325 COILED COLL (POTENTIAL).
 FT DOMAIN 3573 3642 COILED COLL (POTENTIAL).
 FT NP_BIND 154 161 ATP (POTENTIAL).
 FT NP_BIND 1852 1859 ATP (POTENTIAL).
 FT NP_BIND 2133 2140 ATP (POTENTIAL).
 FT NP_BIND 2460 2467 ATP (POTENTIAL).
 FT NP_BIND 2805 2812 ATP (POTENTIAL).
 SQ SEQUENCE 4466 AA; 511772 MW; C465C5C6C60325D CRC64;

Query Match 2.5%; Score 101; DB 1; Length 4466;

Best Local Similarity 19.5%; Pred. No. 89; Mismatches 289; Indels 222; Gaps 39;

Matches 147; Conservative 95; Mismatches 289; Indels 222; Gaps 39;

QY 37 RGLYTGAFGATASHSKS-----LHVPFMGLBGLRMCCFRKLMMTQRMGTGGRDVP 91
 DB 1670 KSTVRSQFADAVSYEEKPRRQWMLDYPAQV-----ALATTQVWTFTE----- 1712

QY 92 LETQMLIES--KSESTDGNSPII-YTVLLPLEGQFRAVLQGNKNEIEICSGDNA 148
 DB 1713 VNISFARLEEGHNSMKDYNNKKQILQNLTLGLIGLK---LTKDRQKIMTIC----- 1762

QY 149 VETDGLHNVYVHNGTNEFEYLQAVKAEKNQOTLHREKKRLPSCLDLFG--WCT 203
 DB 1763 -----TIDVHAR-----DVVAMMVLKKYDSQAF-----QWLSQLRHMAD 1798

QY 204 WDAF-YTDTAEGVEEGLKSLDSOGTFRPLI--IDGWOQIENKADATEC----- 252
 DB 1799 DDKHCVANICDAQFQYSEYEL--GNPRLVITPLTD-----RCYTITLQSL 1842

QY 253 -LYQGAQFATRLYGIKRNTKFGOKLQNEQMSGLKHLVHGAKOHNNK---NVYVHNL 308
 DB 1843 HLVMGSAFGAGPAGTGKTEYTK-----DLGRALGIMVYVNGCSBQMDYKSGNITYGLAQ 1896


```

OY 309 AGWGVKPAATGMEHYDATALAYVQSPGLGNPDIVMSLAHVGLGHPKRVENFN 368
DB 1897 TGAMG-----CFDEFNLSIVELSVAVQKCODAIR-----DKKERNFMG 1939
OY 369 ELHAYLASGVGVKVDVONITETGAGHGVSLSTRSYHALLASIASNFTDNCIACM 428
DB 1940 EELSLIPSVG-----IFITMMPGYAGRTLEPENL-KALFRPCAMVVPDFELICEI 1988
OY 429 CHTNDLYSAKOTA-----IVRASDFYRDRASHIITHISSAYNSLFGEMQ 477
DB 1989 MVAEGLFELARLLARKFTITLYLCKELLISKODHY--DMGLRAIKSVLVVAGSLKRGDQR 2046
OY 478 PWDMEHSHIPADYHAAARAGCPIYVS-----DKPGENFDLTK----- 519
DB 2047 PEDQVLT--MRALDRFNVPKIVSDPTFVEMGLIGDLFPALDVRRRRDDEKVKVOSTIDL 2104
OY 520 KLVLPDGSYLK-AQLPG-RPTDLSLFV--DPARDRTSLIKTNLKKSGVGVENCQAG 575
DB 2105 KLAQEDSEFLKVVQLEELLAVRHVSFVIGMAGTKSOYLKV--LNKT----- 2149
OY 576 WCKIEKTRIHDPSTPGLT-----ASVCASDVLLITQVAGAEWL-- 614
DB 2150 YENMKRKPFVDLNRKATYNDLFGIINPATREKMDGLFSVLRMSNITH--DQKMWTL 2208
OY 615 -GD-----TIVAYRSGEVIRLPKGVSIPTLVK-LEFELHPCPJOELAP-SISFPA 664
DB 2209 DDDIDPMNIESINTVMDNKNKVLTLASNERIPTLPSMRILEFISH--LKTATPAVVSAG 2265
OY 665 IGLDMFNMG-----GAVEQVEIHNRATKTI 691
DB 2266 ILIYNPSDLGNVPYITSWIDREVOSERANLTI 2298

RESULT 15
ID DYNC_TRIGGER STANDARD: PRT: 4466 AA.
AC P23098:
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dynein beta chain, ciliary.
OS Tripneustes gratilla (Hawalean sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidae; Euechinoidae; Echinacea; Temnopleuroidea; Tokopneustidae; Tripneustes.
OX NCBI_TaxID:7673:
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE-Blastula;
RX MEDLINE-91326103; PubMed-1830927;
RA Gibbons I.R., Gibbons B.H., Mocz G., Asai D.J.;
RT "Multiple nucleotide-binding sites in the sequence of dynein beta heavy chain.";
RL Nature 352:640-643(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92020893; PubMed-1833761;
RA Gibbons I.R., Asai D.J., Chung N.S., Dolecki G.J., Mocz G.,
RA Phillipson C.A., Ren H., Tang W.Y., Gibbons B.H.;
RT "A PCR procedure to determine the sequence of large polypeptides by rapid walking through a cDNA library.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8563-8567(1991).
CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES. DINEIN HAS ATPASE ACTIVITY.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA), THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59603; CAA42170.1; -.
DR PIR: S17653; S17653.
DR InterPro: IPR004273; Dynein_heavy.
DR Pfam: Pf03028; Dynein_heavy; 1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
KW Coiled coil.
FT DOMAIN 627 643 COILED COIL (POTENTIAL).
FT DOMAIN 733 805 COILED COIL (POTENTIAL).
FT DOMAIN 1036 1056 COILED COIL (POTENTIAL).
FT DOMAIN 1306 1337 COILED COIL (POTENTIAL).
FT DOMAIN 1443 1468 COILED COIL (POTENTIAL).
FT DOMAIN 1950 1978 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 3033 3092 COILED COIL (POTENTIAL).
FT DOMAIN 3263 3325 COILED COIL (POTENTIAL).
FT DOMAIN 3573 3642 COILED COIL (POTENTIAL).
FT NP_BIND 154 161 ATP (POTENTIAL).
FT NP_BIND 1852 1859 ATP (POTENTIAL).
FT NP_BIND 2133 2140 ATP (POTENTIAL).
FT NP_BIND 2460 2467 ATP (POTENTIAL).
FT NP_BIND 2805 2812 ATP (POTENTIAL).
FT VARIANT 611 615 MISSING.
FT VARIANT 3356 3358 LPS -> LITGNFCCFPMAG.
SQ SEQUENCE 4466 AA; 511771 MW; 2A695BF8336911E CRC64;

Query Match 2.5%; Score 101; DB 1; Length 4466;
Best Local Similarity 19.3%; Pred. No. 89;
Matches 150; Conservative 94; Mismatches 264; Indels 286; Gaps 42;

OY 37 RGLVTGAFVATYSHSKS-----LHVPKGVLEGLRPMCCFRKLLMMTORMTCGRDVP 91
DB 1670 RSTYRSQFADAVVSYEKPREQWLDYPAQV-----ALATQWMTTEVNISFAR-- 1719
OY 92 LETQFMLESKESETPDQENPPI--TYVLLPILLEGOFRAVLOQNDKNEIEICLESQNAV 149
DB 1720 -----LEBEHNSMKDYNNKQIQQLNTLIGLIGK--LTKDROKKIMTIC----- 1762
OY 150 ETDGLHMYVMHAGTNPFEVINYQAVAKENKMTFLRREKRLPSCLDMFG--WCTW 204
DB 1763 -----TIDVHAR--DVAVMMVLKVDNAQAF-----QWLSQRLHNMAD 1799
OY 205 DAF-YTIDYTAEGVEBGLKSLSSQGTTPRPLI--IDGMOQIENKAKADATPC----- 252
DB 1800 DKHCYANICDAQPKSYEYL--GNTPLVLTPLTD-----RCYITLTQSLH 1843
OY 253 LVQGAQFATRLTGKIKENTKFKKLLONNEOMSGKLHIVGAKOHNVK--NVYVHMLA 309
DB 1844 LVMSGAPAGAGKGTETTK-----DLGRALGIMYVYVNCSEQMDYKSCGNITVGLSOT 1897
OY 310 GYGVGVKPAATGMEHYDATALAYVQSPGLGNPDIVMSLAHVGLGHPKRVENFNE 369
DB 1898 GAGM-----CFDEFNLSIVELSVAVQKCODAIR-----DKKERNFMGE 1940
OY 370 LHAAYLASGVGVKVDVONITETGAGHGVSLSTRSYHALLASIASNFTDNCIACM 413
DB 1941 EELSLIPSVG-----IFITMMPGYAGRTLEPENL-KALFRPCAMVVPDFELIC 1986
OY 414 -----SIASNFTDNCIACM-----CMC-----HNTDLYSAKOTATAYRAS--DDFY 451
DB 1987 EIMLVAGLEGLARLLARKFTITLYLCKELLISKODHYDMGLRAIKSVLVVAGSLKRGDQR 2046
OY 452 PRDP-----ASHIITHISSAYNS--LFLGFMOPDMDMHSHPADYHAAARAGCPIY 505
DB 2047 PEDQVLTMRALDRFNVPKIVSDPTFVEMG-----LIGDLFPALDV----- 2085
OY 506 VSDKPNHNFDLLK-----KVLDPDGSYLK-AQLPG-RPTDLSLFV--DPARDRTS 552
DB 2086 -----FRRRMDPEKVKVOSTIDLKLAQEDSEFLKVVQLEELLAVRHVSFVIGMAGTKSQ 2141

```

```

QY 553 LKTMNLKCGVGVVFNCGAGWCIEKKRIHDTSPGTLT----- 594
Db 2142 VLKV--LNKT-----YSNMKKRVLDLNPKAVTNDELFGIINPATREKDG 2186
QY 595 -ASVCASDVDLITQVAGAEMT---GD-----TIVYVRSGEVIRLPKGVSIPTVLKV- 642
Db 2187 LFSYIMRDMNITH-DGPKWIVLDGDIDPMMIESLNTVMMDNKKVLTLASNERIPLTFPSMR 2245
QY 643 LEFELFHPCPIQETAP-SISFAIGLDMENTG-----GAVEQVEIHNRAATKI 691
Db 2246 LFFEISH--LKTATPATVSRAGILYINPSDLGWNPIVTSWIDTREVQSERANLTI 2298

```

Search completed: June 4, 2002, 09:10:09
 Job time: 236 sec

[illegible]

QY	1	MVTPEKISVNDGKLVYHGKTIITLGVDPDNNVLTPLSGSGRLVTCAGFATASHSKSLHVFPM	60
	1		
	1	MTVAGAGISTDSDLVYLGHVRLHGVPEENVLTPLPAGMNLIDAFIGVTSIDTQSHRVSFL	60
QY	61	GVLEGIREFMCCRFKLMWMTQRMGCGRQVPLETOFMLESKE-SETQGENSPILTYVTL	11
QY	61	GKLEDRFCVCFRFLMMWTQRMGNGKNGKEIPETOFLIVEANQSGDLGGRODSSSYVFL	12
QY	120	PLLEGGFRAVLQDGNKEIEICLESGBNAVETDGLHMVYIMAGTNPREVINOAKAV--	17
QY	121	PLIEGGFRAVLQGNANEIEICLESGEDPTVDQFEGSHLVFAAGSDPDVDTLKAVKYMF	18
QY	178	-EKHMOTLHRRKKRLPSCDLDFGFCMTDAFTYDTYACVGEGLK-----SL	22
QY	181	LKSQKLSICLPNEFKRPMDLNMFNGMCTDAFTYNTADVQKESNCDLTRPALILCSL	24
QY	224	SOGGTPRPRFLIDDDMOQOIEENKAKATECIVQEGAFQPTRLTGIENTKFOKTLONNDOM	28
QY	241	KAGVGTPKFVILLIDGMQSY-GMDEISVEFNMNANFNLRLTHIKENKFKQDQEGHRV	29
QY	284	S-----GLKHLVGAQKHNNKNTVYMHALLAGVGGVCPKPAATGMEHYDTALAPVQSGVL	33
QY	300	DDPSLSLGVITVDIKSNNSLKVVYWAHITGYMGVKKPGVSGMEHYESKVALPVSPGVM	35
QY	340	GNQPIVDLSLAVHGLGVHPKKVPEFNYNELHAYLASGVGVGVVDVQNIETLGAGHG	39
QY	360	SENGCCLESTIKNGLVNPBKVFSPFNDLHSLYASVGVGVVDVQNIETLGAGHG	41
QY	400	RVSLTRSYHHALEASIAENFTDNGCIACMCNHTGILYSAKOTAYIRASDDFPPDPASHT	45
QY	420	RYKLAKKYHOALEASISNPNPDNGIISCMNHTGILYSAKTAYIRASDDPFPDPASHT	47
QY	460	IIHSSVAANSFLGFEKPPDMDMHSLSHPADYIAAARAIGCPTIYSDKRGNNHFDLK	51
QY	480	IHIAVAANTLFLGFEKPPDMDMHSLSHPMAEYHAAARAAGGCAIYSDKRGQHDNMLR	53
QY	520	KLVLEDSVLSRAOLPGRPTRSLFVDPARDTSLKLTMLNLCGSGVGVFNCQAGACKI	57
QY	540	KLVLRDGSILRAKPLRGRT--	55
QY	560	EKKTRIHDTSPGTLTASVACASDVDLITQVAGAEWLGDITVYARSGEYIRLPKGSIPPT	63
QY	560	-----ELVYLPKDTSLPVT	57
QY	640	LKVLEFELHFCPIQETLAPISFPAIGLDMFNTEGAVEQVEYHNRAATYITALVSGRG	69
QY	574	LMPREYEFVYPAKESVDSGSKFAPVGMETMFGAIVSLRYRDEGCTKFEVREKMLRSG	63
QY	700	REGVYSS-QRPLKCVVGAETDE 721	
QY	634	LVGVYSSVRRPRSVTVDSDDVEY 656	
RESULT	7		
Q9M442			
Q9M442	PRELIMINARY;	PRT;	386 AA.
AC	Q9M442;		
DT	01-OCN-2000 (TREMblrel_15, Created)		
DT	01-OCN-2000 (TREMblrel_15, last sequence update)		
DT	01-JUN-2001 (TREMblrel_17, last annotation update)		
DE	PUTATIVE IMBIBITION PROTEIN (FARGMENT).		
OS	Cleier arietinum (Chickpea) (Farfango).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eumossids I; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.		
NCBI	taxid=3827;		
FN	[1]		
RP	SEQUENCE FROM N.A.		
RA	STRAIN=CV. CASTLEANA. TISSUE=ETIOLATED OSMOTIC STRESSED EPICOTYLS;		
RA	Dopico B., Romo S., Labrador E.;		
RT	A putative imbibition protein is expressed in chickpea epicotyls.";		
TL	Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.		

DR	EMBL:	AJ271668;	CAB71135.1;	-.
FT	NON_TER	1	1	
SQ	SEQUENCE	386 AA;	42180 MM;	357AF9BC8D71650 CRC64;
Query Match		43.3%;	Score 1732;	DB 10; Length 386;
Best Local Similarity		84.7%;	Pred. No. 3	3e-135;
Matches 327;		Conservative 13;	Mismatches 22;	Indels 24; Gaps 1;
OY	389	IIETLAGHGGGRVSLTRSYHHAALEASIASNFTDNGCIACMCHNTDGLYSAKOTATVRASD	448	
Db	1	IIETLAGHGGGRVSLTRSYHHAALEASIASRNFADNGCIACMCHNTDGLYSAKOTATVRASD	60	
OY	449	DEPRDPASHTHISSVAANSFLGFMOPDMFMSLHPADHYAAAAGCPTIYSD	508	
Db	61	DPRPDPASHTHISSVAANSFLGFMOPDMFMSLHPAEYAHAARAIGCPTIYSD	120	
OY	509	KPGNNFDLLKLVLPDGSVLRALQRLRPDSLFEDPARDRSTSLKINLNKSGGVVY	568	
Db	121	KPGNNFDLLKLVLPDGSVLRALQRLRPDRTRCLFDPARDRTSLKINNMKKCTGVVY	180	
OY	569	FNCQGAGWCKIEKTRIHDTPSQTITLASVCASDVLLITVGAGEWLGDITIYAYSSEVI	628	
Db	181	FNCQGAGWCKVEKTRIHDTSPQTLTSSVASVDQINOVAGEWHGETIYAYRSGEVI	240	
OY	629	RLPKGSIIVTLKVLEFELFHPCPIOEIPISIFAIGILMDPNNGANAEVYEIH-----	683	
Db	241	RLPKGISIVTLKVLEFELFHPCPIOEIPISIFAIGIMDFNNGAEEVEIHKASDN	300	
OY	684	-----NRATKTIALSVGRGHFGFYSSQRPLKCVCVGAETDPND	724	
Db	301	KOLFEGEVASELTTSLSNPRTKATVALVKVSGKFYVSQQHPLQCVAIDSIDFPND	360	
OY	725	SEGTGTTESIPSPSEMYRWSIEIOY	750	
Db	361	SETGTTFSIPVPQEGMYRWSIEIOI	386	
RESULT	8			
OGEND9	ID	PRELIMINARY:	PRT:	783 AA.
AC	OGEND9,			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	RAFIPOISE SYNTHASE PROMEIN.			
OS	Arbidopsis thaliana (mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	Eustoidia II; Brassicales; Brassicaceae; Arabidopsids.			
OX	NCBL_TaxID=3702;			
ON	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=COLDIMBIA.			
RX	MEDLINE=96069011; PubMed=9405377.			
RA	Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,			
RT	Tabata S.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. II.			
RT	Sequence features of the regions of 1,044,062 bp covered by thirteen			
RT	physically assigned pl clones."			
RU	DNA Res. 4:291-300(1997)."			
DR	EMBL; AB006702; BABI1395.1;			
SQ	SEQUENCE 783 AA; 86237 MW; 3C37D1D7871888AC CRC64;			
Query Match	35.6%;	Score 1422.5;	DB 10; Length 783;	
Best Local Similarity 38.7%;		Pred. NO. 4.8e-109;		
Matches 297; Conservative 142;		Mismatches 280;	Indels 49; Gaps 17;	
OY	4	TPKISVNDGKLHVHGKTLTGCVPDNVVL-----PGSGRGIVTGAFAVGATA-SHSKS	54	
Db	19	TEKFLTESILLANGQVYLTDVPAVNVTLTSSPYLVKDQGVLDVDSASFICFNLDGPKS	78	

```

QY 55 LHPKMGVLEBGRMCCFRKLMWMTQMGTCGRDVPLETOFMLESKESET--DGENSPI 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 HHVASIKKIRIRMSIFRFKVMWMTTHWVSGNDRDIENFOIILDDSGSDGPGSGGR 138
QY 114 IYVVLPLLEQFRAVLQGDNDKNEIECLSGDAVETDGLHVMYMHACTNPEVINO 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 PYVLLPLLESGFRSSFOQSGEDDVAVCESGTEVGSERQIVYVHNAHDGPKLKDA 198
QY 174 VKAVERKHQFPLHREKRLPSCLDPMGCTWDAFYTDVTAAGEGLKSLSGGTPPRFL 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 199 MKVIRVHMNTKLEKSPGIVKFCWCTMDAFYLVNPDGVHKGKCLVDGCGPPGLV 258
QY 234 IIDDGMOIENKADATEC---LVDSGAQFATRLTGIKENTKQKLNNEQMS--GLKH 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 LIDDMQSIGHDS--DGIDVEGMNITVAGEQMPCLLKEEENKFKDVSVPSPGDNDVGMKA 317
QY 289 LVHGAQ--QHNVKNVYVHNAALAGVGVKRAATGMENHDTALAVPVQSPGLNQPDIVM 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 318 FVRDLKDEFSTVDIYVHNAALCGVWGLRPAPALP--PSTITRPELSPGKLTMEDLAV 375
QY 348 DSLAVHGLGVHPKKVFNFYNELHAYLASCGVDGVKVDVONIIEFLGAGHGRVSLTRSY 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 DKIIETGIGFASPLAKREFYEGLSHLDNAGIDGVKVDVHILMLQOKYGDGRVDLAKAY 435
QY 408 HHAEASIASNFTDNGCIACMCCHNTDGLY--SAKQTAIVRASDPFYPRDPAS----- 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 FKALTSVNHKNGVGLASMEHCNDFMFLGTEALSLGRVGDWCPTDPSGDPNGTEWLO 495
QY 458 --HTHISVAYNSLFLGEFQOPMDMFSHLHPADYAAARAIGGCGIYVSDRPNHNF 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 GCHMWH---CAVNSLMWGNFIQPDMDMFOSTHPCAEFHASRAISGGIYISDCVGHDF 552
QY 516 DLKRLVLPDGSVLRQAOLPGRPTDSLFPDPAARDTSLIKIMNLKSGGVGVENCQAG 575
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 553 DLKRLVLPNGSLILKCEYALPTRRLRLEDPDLHOGKTMKIMNLKYGIVGAEFCQCG 612
QY 576 WCKIEKTRIHDTSPGTLTASVCAVDLIT-----QVAGAEMLGDTIYVARSSEVRL 630
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 613 WCREFRNRNOCFSECYNVTLTATSPKDVEMNSGSSPISIANVE--EFLALFISQSKLLS 669
QY 631 PGVSLPVLTKLEFELHFCPIQEI--APSISFAIGILDMFNTGGAQVEIHNRAATK 689
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 670 GLNDELTLLEPFKEKELITVSPVVTIEGNSVAFRIGLVNMLNTSGAIRSL--VYN--DE 725
QY 690 TIALSVGRGRGVYSSORPLKCVGGAETDENYDSEGLTFPSIPVS 737
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 726 SVEVGVGAGEFRYVASKKPVSCLLIDGEVEYRGY--EDSMVWQVPWS 771

```

```

RESULT 9
092T62 PRELIMINARY: PRT: 784 AA.
AC 092T62;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE RAFFINOSE SYNTHASE (EC 2.4.1.82).
GN RFS.
OS Cucumis sativus (Cucumber).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
OX NCBI_TaxID=3659;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Ohsumi C., Nozaki J., Kida T.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
KM EMBL: AF073744; AAD02832.1; -.
Transferase: Glycosyltransferase.
SQ SEQUENCE 784 AA; 86920 MW; 3B06AA91F0908933 CRC64;

```

Query Match 34.9%; Score 1394; DB 10; Length 784;
 Best Local Similarity 37.6%; Pred. No 1,1e-106;
 Matches 289; Conservative 143; Mismatches 285; Indels 52; Gaps 15;

```

QY 8 SYNDGKLVYHGKTIILTPGPDVNLTPGSGRGL-----VTGAEVATASHSKSLHVPKGV 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 26 AIDGSDFTYVNGHSFLSDVENVATASPSPYTSIDKSPVSGVCGFDGSDAPSRHVSIGK 85
QY 63 LQGLRPMCCFRKLMWMTQMGTCGRDVPLETOFMLESKESETGENSPITVYVPL 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 86 LMDIRFMSIFRRKVMWMTTHWVSGNDRDIENFOIILDDSGSDGPGSGGR 138
QY 123 EGOFRVAVLQGNKNEIECLSGDAVETDGLHVMYMHACTNPEVINOAVKAEKHM 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 EGFRFSTIDPGDDDFVDVCEGSSKRVYDASRSMILYLAHGDDPRALVKEAMKYRTHLG 198
QY 183 TFLHREKRLPSCLDPMGCTWDAFYTDVTAAGEGLKSLSGGTPPRFLIIDDGMOI 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 199 TFLRLEKTPRPGIVDKFCWCTMDAFYLVHPQGVIEGVHILVDGCGPPGLVILDDGMOI 258
QY 243 ENKAKATECLVQD---GAQFATRLTGIKENTKFO-----KLNNEQMSGLKHLVHGAQ 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 GHDSDEITREGMNVAGBQMPCLLKEFQENKFRDYVNPKATGPBAGCKGKAFITDELK 318
QY 295 -QHNVKNVYVHNAALAGVGVKRAATGMENHDTALAVPVQSPGLNQPDIVMDSLAVH 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 GEERTVEHYVHNAALCGVWGLRPAPALP--EARYIQVLSRGLQMTMEDLAVAKIYVH 376
QY 354 GLGLVHPKKVFNFYNELHAYLASCGVDGVKVDVONIIEFLGAGHGRVSLTRSYHHALEA 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 377 KVLVLPPEKAEEMVEGLHAEKVGIDGVKIDVHILMLCEDYGGVGLAKAYKAMTK 436
QY 414 SIANSFTDNGCIACMCCHNTDGLY--SAKQTAIVRASDPFYPRDPAS-----HTH 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 437 SINKHFRNGVGLASMEHCNDFMFLGTEALSLGVGDDEFCWCTDPSGDPNGTEWLO 496
QY 462 ISSVAYNSLFLGEFQOPMDMFSHLHPADYHAAARAIGGCGIYVSDRPNHNFLLKL 521
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 497 ---CANDSLMWGNFIHPDMDMFOSTHPCAEFHASRAISGGIYVSDVSGKHNFLLKL 553
QY 522 VLPDGSVLRQAOLPGRPTDSLFPDPAARDTSLIKIMNLKSGGVGVENCQAGKWKIR 581
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 554 VLPDGSILRSEYVALPTRDCLFEDPLHNGETMELKIMNLKFTGIVAFNCGGGWCREFR 613
QY 582 KTRIHDTSPGTLTASVCAVDLIT--QVAGAEMLGDTIYVARSSEVRLPFGVSPYT 639
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 614 RNOCFQYSKRVTSKINPKDIEHMSGENPISIGVTFALYTLQAKKLLSKFSQDLDA 673
QY 640 LKVLFEFELHFCPIQEI--IAPSISFAIGILDMFNTGGAQVEIHNRAATKIALSVGR 698
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 674 LDPEFELITVSPYTKLIQSLHAPRIGLVNMLNTSGAISOVDYDDLGS--SVEIGVKG 731
QY 699 GFRGVYSSORPLKCVGGAETDENYDSEGLTFPSIPVSEEMKRSIE 747
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 732 GEMVFPASKKPRACRIDGEDVGFYDQDQ--MVVQVQ-----WPD 771

```

```

RESULT 10
0942T6 PRELIMINARY: PRT: 816 AA.
AC 0942T6;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PUTATIVE RAFFINOSE SYNTHASE.
GN Oryza sativa (Rice).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.

```


RC STRAIN=CV, NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nippobare(Ga3) genomic DNA, chromosome 1, PAC
 clone:00583G08."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP003282; BAB64768.1;
 SQ SEQUENCE 816 AA: 89588 MW: B316EDF3566C5178 CRC64;

Query Match 33.7%; Score 1348.5; DB 10; Length 816;
 Best Local Similarity 37.5%; Pred. No. 7,1e-103;
 Matches 301; Conservative 116; Mismatches 288; Indels 97; Gaps 18;

QY 5 PRISVNDGKLVYHGKTLITGVDPNVLTGSG-----RGLVTGAFVGAATASHKSLHY 57
 DB 26 PRFTLKGRDLAVDGHPELIDVPAIRLTPASTLVNSDVPAALAASTFGFADPAKADHY 85
 QY 58 FMAGVLEGLRFMCCEFRKIMMTOMKGTGCRDVPLETQPMLE---SKSETDGENSPII 114
 DB 86 VPIGKLRDTRFMSIRFKVMTTHVGTNGRDVENETOMIIDQSGTKSSPT---GPRP 141
 QY 115 YTVLLPLEGOFRAVLIQ-GNDKNEIEICLESNDNAVETDQGLAHMYMHAGTNPFEVINO 173
 DB 142 YVLLPIYBGPFRACLESKADYHNVLESSTVRSVSAYVYLAHAGDPEDLVKDA 201
 QY 174 VVAVERHNOTFLHREKKRLPSCLDMFGMCTDAFTDYTAEGVEGLKSLSGGTTPREL 233
 DB 202 MHVVRANHGTFRLEMEKTPRPIVDKFGMCTMDAFYLVKHPGVSVMGVARLADGCGPRELV 261
 QY 234 IIDDGMOOI---ENKAKATEL--VQSGAQFATRLTIKENTKFOKKLQNNQNSGLKH 288
 DB 262 LIDDGMOISICHDDDLGSGAEGEMNTSAGEMPCLIKFOENYKFR-----EYKGMWG 315
 QY 289 LVHGAK-OHHNVKNYVYMHALAGYWGKVPATGMEHDTALAYVPOSGVLGNQPDIVM 347
 DB 316 FYREKKAFPRVEGYVYVNHALCGYWGGLRPGARGLP--PAKVAPRUSPGLQRTMEDIAV 373
 QY 348 DSLAVHGLVHPKKVNFYVYELAHAYLASGVDGVKVDV----- 386
 DB 374 DKIVNNGVLDPRRARRELVEGLSHLQASGIDGVKVDVIRHKYWIILYKIKDNATERFVL 433
 QY 387 -----QNTIELGAGHGGRVSLTSSYHHALEASIASNFTDNGCIACCHNTD- 433
 DB 434 KSEIRKDCVLLQLENVCEYEYGGRELAKAYRAGLTESYRHRHNGVYASMEHCNDF 493
 QY 434 GLYSAKQTAIVRASDDFYRDPAS-----HTHISSAVYNSLFLGEFNOQPDMD 482
 DB 494 MLTGEAVALGRVGDDEFCSTDPDPOSTFWLQCGHNVH---CAYNSLMMGAFTHPQDM 550
 QY 483 FHSIHPADYHAARAIIGGCPITYSDKPGNNINFDLLKLVLPDGSVLAQOLPGRPTRDSL 542
 DB 551 FOSTHPCAFPAASRAVSGGFPVYSDAAGCHDFDLRLRLAPDGTIIRCEERYALPTRDL 610
 QY 543 FVDPARDTSLIKIMNLKRCGGVGVFNCOGAGWCKTEKTRIHDTBPTGLTLAVCASDV 602
 DB 611 FADPLHOKTMLKIMNNKESGVIGARFCOGGMSRERRKMCAGSVPTAASPADV 670
 QY 603 DLITQVAGAEWL-----GDTI-VYAASGEVIRLPKGVSIPTVLKLVLEFELFHPCPTOEI 656
 DB 671 -----EMSHGGGGGDRFAVYFVEARKLOLLRDESEVELLEPTYELLYVAPRAI 721
 QY 657 AP---SISFAIGLDMFNNGANVEQVEIHRKATKITALSVGRGFRGYSSORPLKCV 713
 DB 722 VSPELGIGFAITGLANNMLNGGAVOGFEAARKDGDVAEVAVAGGDMVAVSSARPLCK 781
 QY 714 VGAETDENYDSETGLTTFPSIP 735
 DB 782 VNGDAERKY--EDGIYTVDP 801

AC 09SYJ4;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE PUTATIVE RAFINOSE SYNTHASE OR SEED IMBIBITION PROTEIN.
 GN T7B11.23 OR AT4G01970.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Huang E.N., Nascimento L., de la Bastide M., Habermann K., Vil M.D.,
 RA Preston R.R., Spiegel L.A., See L.H., Shah R., O'Shaughnessy A., Rodriguez M.,
 RA O'Shaughnessy A., Rodriguez M., Sheher M., Swaby I., Schutz K.,
 RA Parnell L.D., Dedhia N.N., McCombie W.R.;
 RT "Arabidopsis thaliana BAC T7B11 from chromosome IV near 10 cm."
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,
 RA Preston R.R., Matero K., Shah R., O'Shaughnessy A., Rodriguez M.,
 RA Sheher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC007138; ABD2659.1;
 DR EMBL: AL161493; CAB80690.1;
 SQ SEQUENCE 807 AA: 90122 MW: 963DCD5A827B338B CRC64;

Query Match 32.7%; Score 1307; DB 10; Length 807;
 Best Local Similarity 37.3%; Pred. No. 1.9e-99;
 Matches 290; Conservative 131; Mismatches 287; Indels 70; Gaps 18;

QY 8 SVNDKLVYHGKTLITGVDPNVLTGSGRLVT-----GAFVGATAS 50
 DB 35 NLESGSLCANDSPPILEDVQNTFTPFSSHSISTDAPILRLRYQANHKSGFLGFTPE 94
 QY 51 HSKSLHFPWGVLEGLRFMCCEFRKIMMTOMKGTGCRDVPLETQMLIESKSETDGEN 110
 DB 95 SPEDRLTNSLGRFEDREFLSLFRFKMMWSTAWIGKSGSDLAQETQVMMLKPEIDS---- 150
 QY 111 SPIITYVLLPLEGOFRAVLOGNDKNEIEICLESNDNAVETDQGLAHMYMHAGTNPFEVY 170
 DB 151 -----YVAIIPTEIGAFRASLTPEKENVILCAESGSTYKESSEFKSIATIHICDNPYNLM 206
 QY 171 NOAVKAVEKHMOTFLHREKKRLPSCLDMFGMCTMDAFYDYTAEGVEGLKSLSGGTPP 230
 DB 207 KEAFSALRYHMMNTFKLLEKKLPKIVDKFGMCTWDACYLTVPAITWGVKEFEDGVCP 266
 QY 231 RFLIIDDGMOOI---ENKAKATELVOGAQFATRLTIGIKENTK---OKKLNNDGM 283
 DB 267 KFYIIDDGMOISIFDDELDKDA-ENVLYGGEOMYARLSEKCKKFRNYKESISGSDV 325
 QY 284 SG-----LKHVHGAKOHNVKNYVYMHALAGYWGKVPAAATGMEHYDTALAYVQSPG 337
 DB 326 SGSGMAAFTKDL---RLRKSISDDIYVMAHLCGAWGVAPR--MMDLKAKVAPRPLSPS 380
 QY 338 VLGNDPDIYMDSLAVHGLVHPKKVNFYVYELAHAYLASGVDGVKVDVQNTIELFLGAGH 397
 DB 381 LGATMDLAVDKVVEAGLIVHPSKAHEFYDSMHSYLASVGTGAKIDVFTLESIAEEH 440
 QY 398 GGRVSLTSSYHHALEASIASNFTDNGCIACMCHNDGLYSA-KQTAIVRASDDFYRDP- 455
 DB 441 GGRVELAKAYYDGLTESMKNKNGTDVYASMOQNEFFELAKQIISIGRVGDDFWMDPY 500
 QY 456 -----ASHTHISSAVYNSLFLGEFMOPDMDMFHSHLPADYHAARAIIGGCPY 505

Dd		501	GDPGCVWYLGCVHMHCHS---	YNSTIMWMQOMIOPMDMEQSDHVCAEYHAASRAALCGBPVY	557
Oy		506	VSDPRG--NNHFDDIKTKIVLPDGSYLRAQLRGPRRDLFTFDPADRPRLKIKMNLNCS		563
Dd		558	LSDHLGRKASHNEFDLTKLAFFDGTITPKCVHYALPFRDLSFNPLPDFKSLIKIFENFKFG		617
Oy		564	GVAVGVENCOGAGWCIEKTRTIRHDTSPGTLTAASCASDV--DLITTVAGAE--WLGDITIY		619
Dd		618	GVIGFCNFQGAGMSPEEHRRFGKRECYTTVSQTGHVSIEMWDQNEAAGSQVTYGGLLI		677
Oy		620	YAIRSGEVYRL-PRKVSIPYTLKVLEFEELFHRCPIQEIAPRS-ISRPAAGILDMFMTGGAY		677
Dd		678	YKOGESEILLFMNSKSEAKITILEPSAFDLISFPATLEYSSGVRAPAPGLINMFNCVTY		737
Oy		678	EDEVETHNRNAARPTIALTSVRGRCRFVSVSGORPLKCVCVGCAETDFNYDSEGTLTTPSIP		735
Dd		738	QDMKY---TGDNSTINVDYKGBGRFMAYSSSNAPVKCYINDKEAFEFMBEETGKLTFVP		792
RESULT		12			
O9SBZ0	ID	09SBZ0	PRELIMINARY;	PRT; 857 AA.	
AC	09SBZ0	01-MAY-2000	(TREMBLrel. 13, Created)		
DT	01-MAY-2000	(TREMBLrel. 13,	Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19,	Last annotation update)		
DE	GALACTINOL-RAFTINOSE	GALACTOSYLTRANSFERASE (EC 2.4.1.67).			
OS	Phaseolus angularis	(Adzuki bean) (Vigna angularis).			
OC	Eukaryota:	Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;			
OC	Spermatophyta: Magnoliophyta;	eudicotyledons: core eudicots: Rosidae;			
OC	eucotids I: Fabales: Fabaceae;	Papilionoideae: Phaseoleae: Vigna.			
OX	NCBI_TaxID=3914;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20117502; PubMed=10652123;				
RA	Peterbauer T., Mucha J., Mayer U., Popp M., Gloessel J., Richter A.;				
RT	"Stachyose synthesis in seeds of adzuki bean (Vigna angularis):				
RT	molecular cloning and functional expression of stachyose synthase,"				
RL	Plant J. 20:509-518(1999).				
DR	EMBL; Y19024; CAB64363.1; -				
KW	Transferase; Glycosyltransferase.				
SO	SEQUENCE 857 AA; 94949 MW; 85248CAB81165679 CRC64;				
	Query Match	32.1%;	Score 1284;	DB 10;	Length 857;
	Best Local Similarity	35.1%;	Pred. No. 1.7e-97;		
	Matches 296;	Conservative 130;	Mismatches 269;	Indels 148;	Gaps 20.
Oy	9 VNDGRIYVHGKTIITGCPNVNYLT-----PGSGRGIVT-----GAFYGATASHS	52			
Dd	23 LSGDKITVAGVVLISHVNPENVTFSSRSSICVPRDAPSLLQRYTAASHKGFLGFES----	78			
Oy	53 KSLHVP-----MGVLEGLRFMCCEFRKLMMWTQMGMTCGRDVPLETQFMLESKESE	105			
Dd	79 ---HVSPSRLINSLSGSRGNPLSFIPRKTYMTSWMVNGSCGDLQMETOWILIEVPERTE	135			
Oy	106 TDGENSPIITYVLLPLLEGOFRAVLQNDKNKEIELCEISGDNAVYEVDQGLHMYYMAGTN	165			
Dd	136 S-----YVVIIPIIEFSFRALHPGSDDHVKTCAESGSTOVRASSFGALATYVAHT	187			
Oy	166 PFEVINQAAVKAAYEKHQTELHRKKRRLPSCLDMFGMCTWDADFYTDTAAGVERGLKLSLQ	225			
Dd	188 PYLMMEAAVSAALRVHLDSFRLLEEKIYVPRYDKFGMCTDAEFLVLYNPVGMHGLDFS	247			
Oy	226 GGPTRPELLIIDGWOOIENKAKDATE---CLVDEGAQFATRLTGLIENTKFFQKKLQ----	278			
Dd	248 GGVAAPFVVVIDDGWQSVNEDDEDPNEDAKNLYLGEGOMATRLRFBEDCKFRKXYQGLLI	307			
Oy	279 -----NNQWSGL-----KHL-----VHGAK	294			
Dd	308 GPNAFSNPETIELLSKGLEAHBLGKQAALIASGGSDLAEIEMLVKYKNEETIDLFGGK	367			

[illegible]

QY 290 VHGAKOHNNKVVYVHMLAGVGVKPAATGMEHYDTALAYVPSGVLYGNQDIYMD 349
 Db 121 VDCYKHHNNKDYVYVHMLAGVGVKPAATGMEHYDTALAYVPSGVLYGNQDIYMD 179
 QY 350 LAHVGGLVHPKRVNFYVNLHAYLASGVYGVKVVYVNIETLGAGHGGRVSTSYH 409
 Db 180 LAHVGGLVHPKRVNFYVNLHAYLASGVYGVKVVYVNIETLGAGHGGRVSTSYH 233
 QY 410 A-----LEASIASNFTDGIACMCHNTDGLYSAKOTATVIRASDDPRPASHTTHISS 464
 Db 234 AATIMLRRLPLVTLNNGCIACMCHNTDGLYSAKOTATVIRASDDPRPASHTTHISS 293
 QY 465 VAYNSFLGEMQPDWDMFHSHPADYHAARAIGCEPIYVSDKPGNNHFDLKLVL 524
 Db 294 VAYNSFLGEMQPDWDMFHSHPADYHAARAIGCEPIYVSDKPGNNHFDLKLVL 353
 QY 525 DGS 527
 Db 354 DGS 356

RESULT 14
 Q93XK2 PRELIMINARY; PRT: 853 AA.
 AC Q93XK2; 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 ST STACHYOSE SYNTHASE (EC 2.4.1.67).
 GN STSL.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
 OX NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. WONDER VON KELDENDON; TISSUE=SEED;
 RA Peterbauer T.;
 RT "Molecular cloning of a cDNA encoding for stachyose synthase from pea
 seeds.";
 RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ311087; CAC38094.1.
 KW Transferrase; Glycosyltransferase.
 SQ SEQUENCE 853 AA; 95890 MW; CB10F18C8D37B38C CRC64;

Query Match 30.7%; Score 1228; DB 10; Length 853;
 Best Local Similarity 34.5%; Pred. No. 7.6e-93;
 Matches 289; Conservative 113; Mismatches 282; Indels 154; Gaps 17;

QY 9 VNDGRLVHGKTLITGVDPNV-----VLTPGSGRGVLT-----GAFVGATAS 50
 Db 21 LSEKFKVKGFLFDVDENVSEFSSISCKSESNAPPISLQKVLANSKGGFGEFHE 80
 QY 51 HSKSLHVRPMVLEGRMRCERFELMMTQRMGTGSDVPLETQFMLEKSESETDEN 110
 Db 81 TFSDRIMNSIGFNGKDLSTIFREKTWSTOWIGKSGSDLOMETOWILIEVETKS---- 136
 QY 111 SPIITVLLPLLEGQFRAVLGNDKNETEICEISGDNVVERDGLAHMYMAGTPEVY 170
 Db 137 -----YVLIITIEKCFRSLPPGFNDHVKITAESGSTVYKSTENSIAVYHSENPYDM 192
 QY 171 NQAVKAVEKHMOTFLHREKRLPSCLDFGWCOTWDAYTDVYAEVGEGLSKLSOGATPP 230
 Db 193 KEAYSARVHNLNSFRLLBEKTIPLNVDFGWCOTWDAYFLVNPIGIFHGLDDFKGATPE 252
 QY 221 RFLIIDDDMOQLEKAKATE---CLVEGQAFATRLTGIENTKFKQ----- 275
 Db 253 RFLIIDDDMOQLEKAKATE---CLVEGQAFATRLTGIENTKFKQ----- 312
 QY 276 -----KLQNNQMS- 284

Db 313 PYDPNNFTDLLKGLTEHKLKKREELSSKSDLAETESIKKVVKPIDDLFGGEDESS 372
 QY 285 -----GKHLVHGAQ-OHNNKVVYVHMLAGVGVKPAATGMEHYDTALAYPV 333
 Db 373 GEKSEMKSEYGLKAPTKDLRTFKGLDDVYVHMLAGVGVKPAATGMEHYDTALAYPV 429
 QY 334 OSPGVLYGNQDIYMDSLAVHGLGVHPKRVNFYVNLHAYLASGVYGVKVVYVNIETL 393
 Db 430 ISPGGLDGMEDLAVVEISKASIGLVHPSQANLELDSMSYLAESGIGVYVHLSLEYV 489
 QY 394 GAGHGGRVSTSYHHALEASIASNFTDNGCIACMCHNTDGLY-SAKOTATVIRASDDPR 452
 Db 490 CDEYGRVDLAKVYVEGLTKSIGVKNFNNGMIASQCHNDFFLGTOISGRYGDDEF 549
 QY 493 RDP-----ASPTHTISSVAYNSLFTGEFMDPDWDMFHSHPADYHAARAIGG 501
 Db 550 QDPNDDPMGSEFWLOGVHMHC-----YNSLMMGMQIPDWMDFQSDHVCAPKHAASRAICG 606
 QY 502 CPIYVSDKPGNNHFDLKLVLDPGSVLRADLPGRPTPDSLFVDPADRTSLKIRMLNK 561
 Db 607 GPIYVSDVNGSHDPLKLLKLVFPGDTIPKCIYFPLPTDCLFKNPDLFDTHTVAKIMFNK 666
 QY 562 CSGVGVFPCOGAGWCKIEKTR-----IHDTSPGTLTASVCASDVDLITQVAGAEW----- 613
 Db 667 YGVVIGAFNCGAGWMDPIQKFRGPEPCYKPIPTV-----HYTEVEMQKE 713
 QY 614 -----LG---DIIYVAYNSGEV-ILPKGVSLPTVLKYLELEFHCPIQETIAPSISFAI 665
 Db 714 ETSHLGKAEVYVYVINOAEELSLMTLKSEPIQFIOPTSTFELYSFVPTVLCGIGIKPAPI 773
 QY 666 GLDDFNTGAVGEVETIHNRATKTIALSVGRGFRGVSSORLKCVCVGAETDENY 723
 Db 774 GLTNFNSGIVIDLETVGNGA-----KIKVGGGSFLAISSEPKKPOLMNGCEVDFEW 827

RESULT 15
 Q9FVW2 PRELIMINARY; PRT: 283 AA.
 AC Q9FVW2; 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 ST SEED IMBIBITION PROTEIN (FRAGMENT).
 GN DIN10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20532769; PubMed=11080291;
 RA Fujiki Y., Ito M., Nishida I., Watanabe A.;
 RT "Multiple signaling pathways in gene expression during sugar
 starvation. Pharmacological analysis of din gene expression in
 suspension-cultured cells of arabidopsis.";
 RL Plant Physiol. 124:1139-1148(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC PubMed=11240919;
 RA Fujiki Y., Yoshikawa Y., Sato T., Inada N., Ito M., Nishida I.,
 Watanabe A.;
 RT "Arabidopsis genes from Arabidopsis thaliana are associated with
 leaf senescence and repressed by sugars.";
 RL Physiol. Plantarum 111:345-352(2001).
 DR EMBL: AF159578; AAG23721.1.
 DR InterPro: IPR000719; Euk_pkinase.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 FT NON_TER 1 1
 FT TER 283 283
 SQ SEQUENCE 283 AA; 31240 MW; 33C43B84E8540301 CRC64;

